DOCUMENT RESUME

ED 260 545 EC 180 511

AUTHOR	Danaher, Joan, Ed.
TITLE	Assessment of Child Progress. Monograph Number 2.
INSTITUTION	North Carolina Univ., Chapel Hill. Technical Assistance Development System.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE	May 85
CONTRACT	300-82-0369
NOTE	61p.
PUB TYPE	Guides - Non-Classroom Use (055) -- Collected Works - General (020)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Disabilities; *Early Childhood Education; *Evaluation Methods; Infants; Program Evaluation; *Student Evaluation

ABSTRACT

Four author-contributed papers examine issues in assessing child progress in early childhood special education. D. Bricker and S. Gumerlock present "A Three-Level Strategy" which features analysis of daily or weekly progress, analysis of progress toward long- and short-term objectives, and analysis of progress toward program goals. C. Dunst follows with a discussion of "Four Developmental Perspectives" (maturational, behavioral, process-oriented, and ecological) and suggests several generally feasible research strategies. N. Johnson-Martin enumerates three "Sources of Difficulty" in assesssing progress in the population: (1) the discrepancy between implicit and explicit goals in programs serving handicapped infants; (2) the diversity of needs of children served in early intervention programs; and (3) the different kinds of training of professionals who staff early intervention programs. In the final paper, P. Strain stresses the concept of "Social Validity" of intervention outcomes, arguing against the use of standardized tests and advocating instead the use of single-subject designs to measure progress. (CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

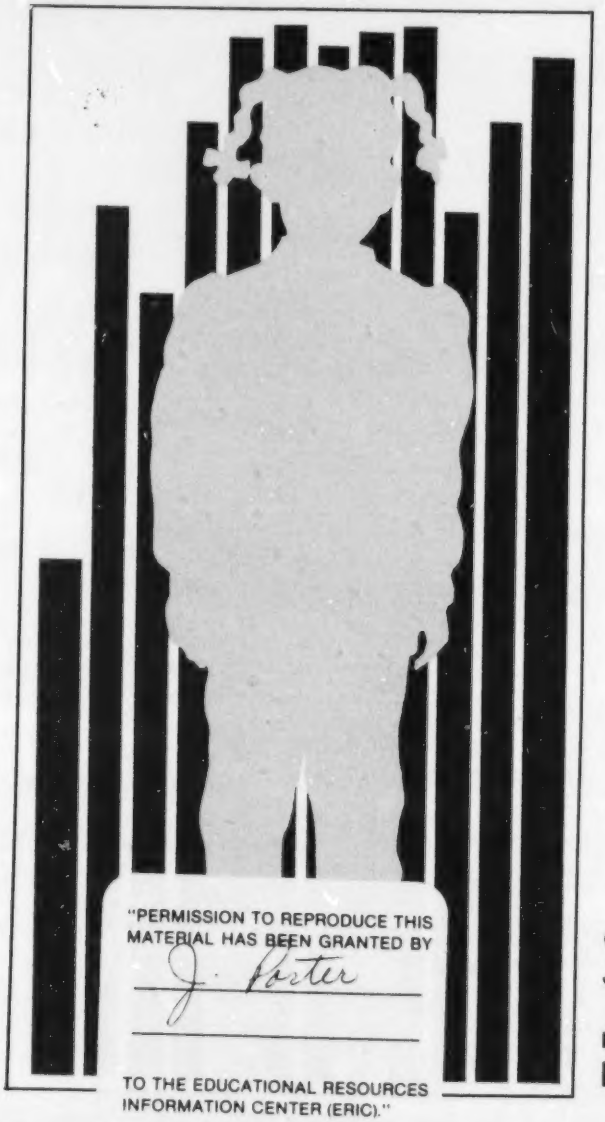
U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED260545

ASSESSMENT OF



"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Porter
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

edited by
Joan Danaher

managing editor:
Daniel Assael

CHILD PROGRESS

EC180511

TADS

Managing Editor: Daniel Assael
Editorial Assistant: Marcia Decker
Production: Pamela Sobek
Printing: University of North Carolina at Chapel Hill

This monograph (number 2) *Assessment of Child Progress*, was produced by the Technical Assistance Development System (TADS), a program of Frank Porter Graham Child Development Center, University of North Carolina at Chapel Hill. TADS is located at 500 NCNB Plaza, Chapel Hill, North Carolina 27514 (tel. 919-962-2001).

This book was prepared pursuant to contract number 300-82-0369 from the Office of Special Education Programs, U.S. Department of Education. Contractees undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions, therefore, do not necessarily represent the Department of Education's position or policy. The contents of this book are presented for information purposes only; no endorsement is made.

Helene Corradino, Project officer to TADS, Office of Special Education Programs U.S. Department of Education

Pascal L. Trohanis, TADS Director and Principal Investigator

May 1985

Contents

Introduction 5

Chapter 1. **A Three-Level Strategy**
by Diane Bricker and Sarah Gumerlock 7

Chapter 2. **Four Developmental Perspectives**
by Carl Dunst 19

Chapter 3. **Sources of Difficulty**
by Nancy Johnson-Martin 39

Chapter 4. **Social Validity**
by Phillip Strain 55

The authors:

Joan Danaher is a technical assistance coordinator at TADS, a technical assistance agency of the U.S. Department of Education's Handicapped Children's Early Education Program (HCEEP). Her areas of expertise and interest include special education, infant curricula and programming, staff development, services to parents, and demonstration and dissemination planning.

Diane Bricker is a professor of special education at the University of Oregon (Eugene) and director of an HCEEP demonstration project there. Her areas of special interest include development of early intervention programs and language programs that emphasize broadly based communication functions.

Carl Dunst is director of an HCEEP demonstration project at Morganton, North Carolina. He is presently researching the influences of social support on families of handicapped and at-risk children. Other interests include the study of factors that contribute to the development of handicapped youngsters and their families.

Sarah Gumerlock is a doctoral student in early childhood special education at the University of Oregon (Eugene). In 1983-84, she was assistant coordinator of an HCEEP demonstration project.

Nancy Johnson-Martin is head of the psychology section of the Division for Disorders of Development and Learning at the University of North Carolina's (Chapel Hill) Child Development Institute. She is also an investigator at that university's Frank Porter Graham Child Development Center. A former director of an HCEEP demonstration project, she has been active in the areas of clinical assessment of infants, curriculum development, and the effects of early intervention on parent-child interaction.

Phillip Strain is an associate professor of psychiatry and special education at the University of Pittsburgh. He is also director of an HCEEP research institute at that university's Western Psychiatric Institute and Clinic. His areas of special interest include the social development of young handicapped children and the development of models for academic and social mainstreaming of normally developing and severely handicapped preschoolers.

Introduction

Early childhood special educators face many problems as they try to assess child progress. Population variability, diverse handicapping conditions, and socioeconomic factors influence the intervention targets; from birth to age 6 years, typical rapid development confounds the effects of the program and the effects of maturation; infants and very young children have limited behavioral repertoires with which to measure change; instruments appropriate for measuring developmental change in severely handicapped or sensory impaired children are either nonexistant or notoriously lacking in technical attributes; and most early intervention programs do not have resources to apply rigorous research designs.

To explore these problems and some solutions, TADS tapped the ideas of several members of the Handicapped Children's Early Education Program. Our contributors are: Phillip Strain, Director of the Early Childhood Research Institute at the University of Pittsburgh; Diane Bricker, Director of the Early Intervention Program, University of Oregon, Eugene, assisted by Sarah Gumerlock, doctoral student in early childhood special education; Carl Dunst, Director of Project SUNRISE, Western Carolina Center, Morganton, North Carolina; and Nancy Johnson-Martin, Investigator with the Carolina Institute for Research in Early Education of the Handicapped, University of North Carolina at Chapel Hill. These experts in the areas of assessment, curricula, and efficacy of early intervention responded to three questions:

> What is child progress? (goals of intervention; appropriate domains for measuring progress)
>
> How can child progress be measured? (types of instruments; purposes of assessment)
>
> What designs or procedures should be used to determine if progress has been made? (problems of design; demonstrating progress as a result of intervention)

The contributors share some points of view within their unique perspectives on assessing child progress. Of the common themes, the writers agree that there is a need to distinguish between assessing child progress for individuals (or groups) and assessing program impact. They also share the idea that measurement of progress should reflect the goals of the program, including benefits to the family of which the child is a member. And they share concerns about the content of assessment (what to measure), and what kind or amount of change constitutes progress. A sampling of the unique perspectives of the individual contributors follows.

Diane Bricker and Sarah Gumerlock propose a three-level strategy for assessing child progress. Each level has its own purpose (child change vs. program impact), referent (child's present performance compared with past performance, child's performance compared with norms, child's performance compared with program goals), and measurement techniques (trial-by-trail data collection for individuals, instruments which reflect long-term goals used across children, standardized instruments). They discuss the problems inherent at each level and suggest some practical solutions.

Carl Dunst focusses on four major developmental perspectives. Each influences the goals of intervention, outcomes to be measured, and techniques of measurement. He emphasizes the need to clearly define the intervention as an independent variable, pose appropriate questions, and employ feasible research strategies to answer the questions.

Nancy Johnson-Martin enumerates three sources of difficulty in assessing progress in young handicapped children. Explicit goals (developmental progress) vs. implicit goals (normal life experience) creates conflicts in determining what to measure. The diverse needs of subgroups of handicapped children dictate that the appropriate domains for intervention will vary, as will the intervention techniques and measures of progress which follow. Further, she says that the professional training of the disciplines involved in early intervention influences the way in which problems are addressed and progress is measured. She describes the types of instruments available and when and how they are most appropriately used.

Phillip Strain stresses the concept of social validity of intervention outcomes. He argues against the use of standardized tests to measure progress of handicapped children. He states that the gains required to produce, for example, a ten-point IQ score increase may be so limited that they are not perceived by significant others in the child's family or community. He describes single-subject designs or, as he prefers to call them, subjects-as-their-own-controls designs and believes they are particularly attractive for determining child progress in early intervention programs.

Chapter 1
A Three-Level Strategy

by Diane Bricker and Sarah Gumerlock

WHAT IS CHILD PROGRESS?

At first blush, this seems a straightforward question that should have a straightforward answer. However, this is not the case. The question of child progress must be related to some benchmark or standard. There must be a source of reference, whether it be the child's performance, the performance of others, or the test itself (Hamilton & Swan, 1981). Progress cannot be determined in the abstract. One must ask: Progress towards what? Measured by what? Once a relationship to a target or goal is established, the following qualifiers must be considered: How much change? What kind of change? Therefore, to answer the question, "What is child progress?" it is first necessary to answer these three questions:

- What content (e.g., benchmarks of development or standards of performance) is to be measured?
- What quantity of change toward acquisition of this content constitutes progress?
- What quality of change towards acquisition of this content constitutes progress?

Measuring Child Progress: Purpose and Content

No accepted set of standards exists which permits universal comparison of children to determine progress. No operationally defined prototype is generally acceptable (Bricker, 1978). For example, all may agree that children should do well in school. But, attempts to define "do well in school" highlight the need for qualifiers. Benchmarks are established or defined in relation to the individual's personal environmental and biological resources, personal values, and the values of the individual's family and culture. (The standards of progress for a biologically normal child are different from those for a child with Down's syndrome or for a child raised in poverty.) Child progress can be determined in relation to school or program goals, goals for specific populations, or goals for a particular child. Since particular comparisons are appropriate for particular purposes, it is important to determine the purpose of measuring child progress before data is collected.

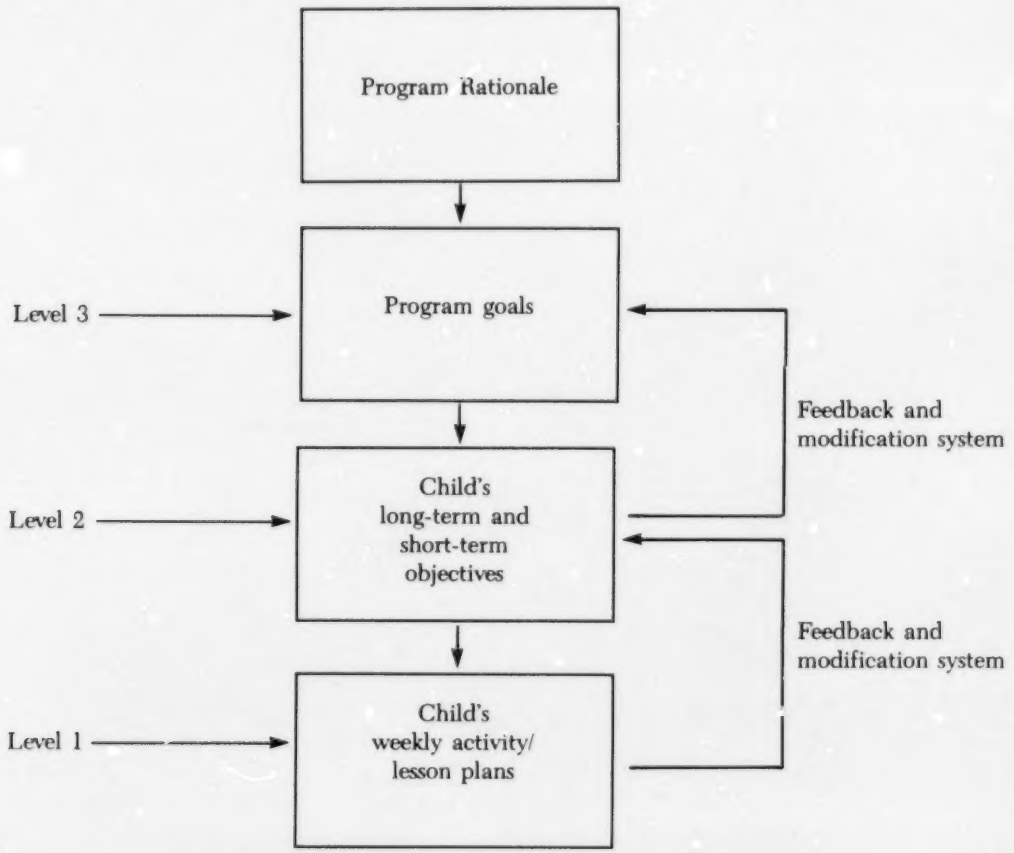
Child Change and Program Impact

Child progress is usually measured in order to evaluate general program impact and to monitor child change. Established program goals can provide guidelines for selecting the content areas for determining child change and program impact. Many early intervention programs lack a cohesive system or plan for determining program goals. This, in turn, affects the selection of goals for children participating in the program. Program goals should provide the broad benchmarks of development or standards of performance against which child progress can be measured.

An organizational framework of three independent, interrelated levels of evaluation can help establish the content areas for measuring child progress. Figure 1 shows one such framework. The greater the consistency between the three levels shown in Figure 1, the more efficient, appropriate, and useful will be the selected areas for measuring child change and program impact.

Figure 1

Three-Level Framework to Direct Program Evaluation



The selection of program goals should be predicated upon the underlying philosophy or rationale that sets the general structure and atmosphere for the program. A frequent rationale for early intervention is that it helps the child grow and develop in important behavioral areas and helps the family accept the handicapped child. Based on this rationale, a set of related program goals might include enhancing: a) children's acquisition and use of sensorimotor skills, b) children's acquisition and use of social skills, c) children's acquisition and use of communication skills, d) children's acquisition and use of motor skills, e) positive family interaction, and f) parent positive attitudes about their children. Such program goals (Figure 1, Level 3) form a content framework from which professional staff and parents can select and develop the individual child's long- and short-term objectives (Level 2). For a specific child, objectives based on program goals might be:

- Long-Term Objective: The child will appropriately use social-communicative signals to indicate labels, requests, greetings, and to gain attention.
- Short-Term Objective (based on the long-term objective): The child wil look at adults, point to objects, and vocalize to gain desired objects, actions, or events.
- Long-Term Objective: The child will initiate appropriate interactions with peers.
  Short-Term Objective: The child will respond appropriately to interactions initiated towards him by peers.

The child's individual weekly activity/lesson plan (Level 1) is based on the child's individual long-and short-term objectives (Level 2). Weekly activity plans for the short-term objectives might be:

- During opening and closing group time, a desired object or event will be withheld while the child is prompted to look at adults, point, and vocalize.
- During snack time, juice and crackers will be placed out of the child's reach and the child will be prompted to make eye contact, reach, and vocalize.
- During small-group activities, peers will be prompted to initiate interaction with the child (e.g., share a crayon, throw a ball), and the child will be prompted to reciprocate (e.g., give peer his or her crayon, roll the ball back).
- During any program activity the child will be immediately reinforced (given desired object, event, person) when he or she spontaneously makes eye contact, reaches, and vocalizes.
- During any program activity, the child will be immediately reinforced (praised, given another turn) for responding appropriately to peer's actions.

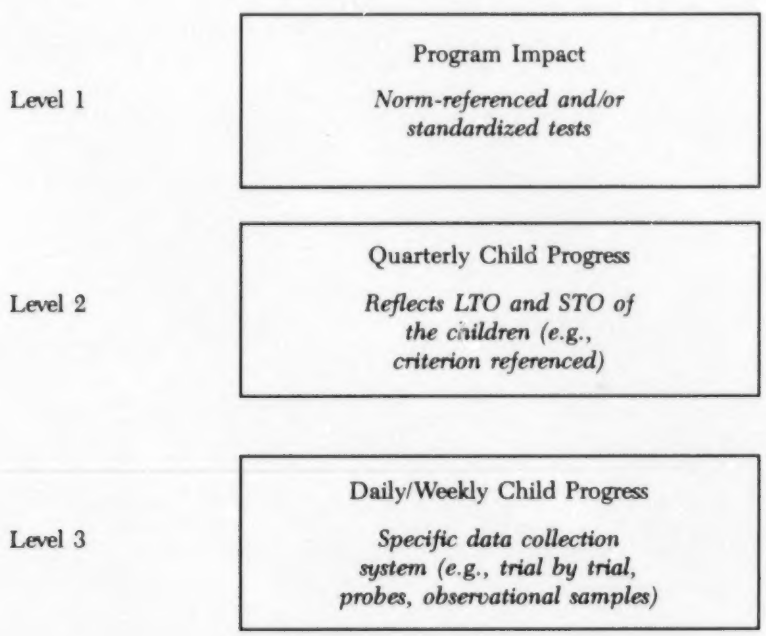
These examples reflect a series of training targets that are global at Level 3, then become more specific at Level 2 and Level 1. This system is useful for selecting the content to measure child progress at each of the three levels of evaluation. The measurement strategy and the comprehensiveness of the measurement target may change at these three levels, but the basic content remains the same. This content consistency provides a set of guidelines which directs measurement of child change and program impact.

HOW CAN PROGRESS BE MEASURED?

Once the purpose and set of content benchmarks or standards against which change will be compared are established, then decisions as to how to measure the child against these established standards must be made. Consider the purposes for the evaluation. To evaluate child change, the child's performance may be compared with the child's previous perf mance. However, to evaluate program impact, the child's performance may need to be compared to some external source (such as a standardized test, a set of norms, or a statistical analysis of change). The three-level framework in Figure 1 that provides a cohesive strategy for selecting content can also provide a useful system for measuring child change and program impact. Figure 2 parallels Figure 1 and illustrates appropriate instruments for measuring progress at each of the three levels.

Figure 2

Three-Level Framework to Direct Selection of Instruments for Measuring Progress


Level 1
Program Impact
Norm-referenced and/or standardized tests
Level 2
Quarterly Child Progress
Reflects LTO and STO of the children (e.g., criterion referenced)
Level 3
Daily/Weekly Child Progress
Specific data collection system (e.g., trial by trial, probes, observational samples)

Level 1: Measuring Progress Towards Daily/Weekly Training Targets

In most cases, the measurement of the child's acquisition and use of specific training targets will have to be conducted without the benefit of a standardized or criterion-referenced instrument. Such intruments tend to target more general response classes than are or should be reflected in par-

ticular training targets for children. To measure child change related to particular training targets, a specific data collection format (e.g., trial by trial, probes, observational sample) appropriate to the training target will need to be devised. Since particular formats will be used with particular targets, measurement approaches probably will vary across training objectives. For example, data collected on the child's use of vocalizations and gestures might be monitored through a weekly observation of the child at play. Monitoring the child's responses to his or her mother's vocal imitation might be done by conducting daily trial probes.

The amount and type of data collected should be determined by the severity of the child's disability, the specific training objective, and program resources. Collection of information on child progress can be done with the instructional program. The interventionist can record the child's behavior and instruct a new response simultaneously if the data collection system has been properly devised to consider the nature of the instruction, the child, and the demands on the interventionists. The strategy should balance the program resources against program demands -- monitoring child progress is only one demand.

The Early Intervention Program (EIP) at the Center on Human Development, University of Oregon (Bricker, Bailey, & McDonnell, 1984) used a strategy involving a grid format to record data. This system attempts to assess the impact of intervention using a group format with the training activities/lesson plans integrated into daily, ongoing activities. The data collection grid is a flexible format that allows the intervener to target several different skills for each child across activity groups (see Figure 3). The grid format accommodates individualization of antecedents, responses, and consequences or error corrections; type and frequency of data collection and dispersion of instruction; and data collection across settings, adults, and peer groups. The antecedents, responses, and consequences written on the data grids are a shortened version taken from the individual child's program plan (Figure 4).

Level 2: Measuring Progress Toward Long- and Short-Term Objectives

The information acquired on child change in terms of specific targets is essential to the formulation and implementation of a sound intervention program. The data collected on child progress toward long- and short-term objectives can also be used to measure child change. And, these data can be used to evaluate program impact if the nature of the data can be aggregated as group data for meaningful analyses. To measure progress toward long- and short-term objectives, published instruments with norms or reliability and validity information may be useful. Selection of an instrument should be based on these factors:

- The test reflects the long- and short-term objectives selected for the children.
- The test reflects the program's goals.
- The test can be administered and scored without much difficulty.
- The test's results are interpretable and useful.

For example, the Revised Gesell Developmental Schedules (Knobloch, Stevens, & Malone, 1980) may be appropriate for measuring achievement for a

Figure 3

Sample Data Grid

	Theresa	Michael	Stephen		KEY
Communication	A: wants, needs, ideas R: sign and/or vocalize C: + receive object, praise - repeat model (phys. assistance to sign) 1 2 3 4 5	A: desired object R: vocalize C: + receive object - model 1 2 3 4 5	A: teacher model of word R: word approximation C: + praise - repeat model ball 1 2 3 4 5 book 1 2 3 4 5 box 1 2 3 4 5		A = antecedent R = response C = consequence (error correction) 1 - 5 = number of trials
Social	A: peer offers toy (prompted by teacher) R: looks at peer and takes object C: + praise from peer - physical assistance 1 2 3 4 5				
Cognitive		A: verbal direction R: follow appropriately C: + praise - physical assistance "come here" 1 2 3 "sit down" 1 2 3			
Gross Motor			A. bell ringing; time to transition to next group R: pulls to standing & walks with one hand held to next group C: + plays with objects at next group - physical assistance 1 2 3 4 5		

BEST COPY AVAILABLE

Figure 4

Sample Program Plan

Domain: Communication

Decision Rule: less than 50% for two days (change plan)

Client/Child: Theresa **Student:** _____

Type: _____ **Group** X **Individual** _____ **Home**

Long Term Goal: Theresa will increase over baseline her spontaneous and appropriate use of signs accompanied with vocalizations to communicate her wants, needs and ideas.

Short Term Objective #1: Theresa will imitate an adult model of a sign with vocalization with 90% accuracy over 2 consecutive days.

Program Steps: 1. A: adult verbal and gestural model

R: Theresa signs and vocalizes

C: + give requested item and social praise

– repeat model with physical assistance to sign

Short Term Objective #2: _____

Program Steps: _____

program with a developmental philosophy in which program staff intervene across adaptive behavior, gross-motor, fine-motor, language, and personal-social areas. However, a program concentrating on the inclusion of parents in the assessment process and without the resources to hire a psychologist or a qualified developmental specialist may find the Developmental Profile II II (Alpern & Shearer, 1980) more appropriate.

If an instrument can be found that meets, or can be modified to meet, program-relevant criteria, serious consideration should be given to the test's adoption. A published or broadly disseminated instrument may have norms or psychometric information not available for most "homemade" instruments. A more widely used instrument also may have been developed with more care, thus contributing to the instrument's face validity and usefulness. And, widely used instruments may permit meaningful comparisons across children, groups, and programs. Since homemade tests are often poorly developed and lack data on their psychometric properties, results can be uncertain and suspect. However, for populations that are severely multiply impaired, program staff may be required to develop instruments or make major modifications in existing measures.

Level 3: Measuring Progress Toward Program Goals

Assessing attainment of program goals can be used to monitor child change. Though, measurement of this type generally is more useful for monitoring the more generic impact of the program on participating children and comparing, when possible, different approaches and programs. For level of assessment of progress and impact, adoption of a norm-referenced or a standardized test is the most desirable choice. The interventionist must decide the types of child progress goals for all children or subgroups of children. These goals may be: 1) normal development in all areas, 2) normal development in several areas, 3) any development in all areas, or 4) any development in some areas (May, 1979). If the program goal is normality in the sensorimotor domain, then the Bayley Scales of Infant Development (Bayley, 1969) may be an appropriate instrument to select. Because of their long-term summative nature, standardized tests are a wise choice when evaluating program goals and outcomes (Green, Ginsburg, & Hyman, 1974). However, programmatic evaluation poses problems similar to problems encountered with individual assessment. These problems include wide population variability, unsuitable instruments, and inappropriate design and analytic strategies (Bricker & Littman, 1982). Interventionists should be aware of the limitations of norm-referenced tests and appropriately qualify results (see Garwood, 1982; Ramey & Wasik, 1982).

It is important to establish criteria for selecting an appropriate test. The test generally should have the following characteristics:

- appropriate for the population, though the population on which the test was normed may be different
- administered and interpreted within the resources of the program
- results useable, for but not necessarily limited to, program evaluation
- sensitive enough to detect change in children's performance
- in reasonable agreement with program goals and objectives (Hamilton & Swan, 1981).

WHAT DESIGNS OR PROCEDURES SHOULD BE USED TO DETERMINE PROGRESS?

Again, it is useful to use the three-level evaluation framework in Figure 1 to examine appropriate designs and analytical procedures. Designs and analyses should vary according to the purpose, content, and level at which child progress is being monitored or program impact is being assessed.

Level 1: Analysis of Daily or Weekly Progress

At this level, the child's progress towards the acquisition and use of specific behaviors is of primary interest. Some form of daily or weekly information should be collected and then systematically compared with data collected previously. Data must be acquired in a consistent manner if the comparisons are to be valid. A number of different strategies can be used. For example, the number of objectives reached within a given period of time can be indicated. Data can be plotted on individual graphs to illustrate percent, proportion, frequency, or rate change over time. A single-subject design is probably most appropriate for this level of analysis. Programs should attempt to demonstrate functional relationships between the instructional programs and changes in the child's responses or patterns of behavior. If resources needed to carry out effective reversal or multiple baseline procedures are unavailable, programs should select a pragmatic way to examine and display the daily or weekly data to monitor child progress and to make sound educational decisions.

Level 2: Analysis of Progress Toward Long- and Short-Term Objectives

The analysis described above (for Level 1) can be used at Level 2. If the children can be properly assigned to a group or subgroup, a group design approach would also be appropriate. Program staff have at least two general options for group analysis. The first involves a comparison of predicted progress and actual progress within the program. The predicted progress can be statistically compared to actual progress in several ways, including reference to existing norms, correlation between performance on pretest and posttest, and correlations between posttest of one year and posttest of the next year. Instead of a statistical comparison, expectancies or timelines for attainment of short-and long-term goals can be established. In this way, the child's actual progress can be plotted against the preset timelines for each long-and short-term target. These approaches require assumptions of linear growth that are questioned in the professional literature.

The second alternative requires the comparison of children in a particular intervention program and children in other programs. This design has these problems: comparability of children across programs is difficult to ensure; critical program dimensions (e.g., staff) may vary; and assessments may favor one group over another (e.g., communication assessment that requires only an oral response may unduly penalize children in programs where alternative communication strategies are encouraged).

Currently, no perfect strategy exists for measuring child progress with groups of handicapped children (Bricker & Sheehan, 1981). Programs must evaluate their goals and resources and then select the design that offers the best compromise for their particular population.

Level 3: Analysis of Progress Toward Program Goals

The analytic designs discussed for Level 2 are equally appropriate for this level of analysis. At Level 3, some form of comparison seems mandatory if program impact is to be implied. A number of serious barriers exist when evaluating progress toward program goals. These problems include wide population variability, unsuitable instruments, and inappropriate designs and analytic strategies.

Early intervention programs serve children with impairments that range widely in nature and degree. Also, a wide range exists in the educational and socioeconomic characteristics of families involved in these programs. Often, this diversity allows only extremely small numbers of children to be meaningfully grouped. Heterogeneity affects the measures and designs that are applicable for examining program impact (Lewis & Wehren, 1982).

Existing instruments may be inppropriate for certain populations of handicapped children for at least two reasons. First, selected items may be completely inappropriate for children with certain disabilities. Second, for more severely impaired children, the developmental space between items may not be sufficient to reflect any change or growth.

Most programs cannot randomly assign children to experimental and control groups. Further, nonintervention controls are generally not possible for ethical reasons. The problems with the design options discussed in Level 2 are also present in evaluating progress toward program goals (Bricker, Sheehan, & Littman, 1981).

Barriers that face interventionists as they try to document program impact are significant and underlined with another pervasive difficulty. The majority of early intervention programs do not have enough resources to conduct elaborate and controlled comparisons of program impact. This reality does not excuse interventionists who do not try to document program impact. Neither can poorly conceived and executed measurement plans be condoned. Rather, limited resources should require: 1) thoughtful compromise, 2) extensive planning to deploy limited resources most effectively, and 3) dedication to searching for acceptable alternatives to present designs and analytical procedures.

Summary

Attempts to determine child change and program impact present enormous problems. Currently, the best we can do is to adopt those compromises that produce the greatest benefits. Limitations continue to exist in measuring child change and the impact of early intervention programs, though with the development of the field, appropriate solutions to evaluation problems are forthcoming.

REFERENCES

Alpern, G., & Shearer, M. (1980). The Developmental Profile II, (rev. ed). Aspen, Colorado: Psychological Developmental Publications.

Bayley, N. (1969). Bayley Scales of Infant Development. New York City: Psychological Corporation.

Bricker, D. (1978). Early Intervention: The Criteria of Success. Allied Health and Behavioral Sciences Journal, 1, 567-582.

Bricker, D., Bailey, E. J., & McDonnell, A. (1984). Early Intervention Program. Final Report Submitted to the Handicapped Children's Early Education Programs, U.S. Department of Education.

Bricker, D., & Sheehan, R. (1981). Effectiveness of an Early Intervention Program as Indexed by Child Change. Journal of the Division for Early Childhood, 4, 11-27.

Bricker, D., Sheehan, R., & Littman, D. (1981). Early Intervention: A Plan for Evaluating Program Impact. Seattle: WESTAR Publications.

Garwood, S. (1982). (Mis)use of Developmental Scales in Program Evaluation. Topics in Early Childhood Special Education, 1(4), 61-69.

Green, D., Ginsberg, N., & Hyman, H. (1974). The Nature and Uses of Criterion and Norm-Referenced Achievement Tests. Monterey: McGraw-Hill.

Hamilton, J., & Swan, W. (1981). Measurement References in the Assessment of Preschool Handicapped Children. Topics in Early Childhood Special Education, 1(2), 41-48.

Knobloch, H., Stevens, F., & Malone, A. (1980). Manual of Developmental Diagnosis: The Administration and Interpretation of the Revised Gesell and Armatruda Developmental and Neurologic Examination. Hagerstown: Harper & Row.

Lewis, M., & Wehren, A. (1983). The Central Tendency in Study of the Handicapped Child. In D. Bricker (Ed.), Intervention with At-Risk and Handicapped Infants. Baltimore, Maryland: University Park Press.

Ramey, C., Campbell, F., & Wasik, B. (1982). Use of Standardized Tests to Evaluate Early Childhood Special Education Programs. Topics in Early Childhood Special Education, 1(4), 51-60.

Sheehan, R. (1979). Measuring Child Progress: Large Group Design and Norm-Referenced Alternatives. In M. J. May (Ed.), Evaluating Handicapped Children's Early Education Programs. Seattle: WESTAR Publications.

Chapter 2

Four Developmental Perspectives

by Carl Dunst

WHAT IS CHILD PROGRESS?

A broad definition of child progress is 'actual, inferred, or perceived change in some specific or general dimension of child behavior.' In turn, change and progress imply that development of some sort has occurred. An operational definition of progress and change depends upon one's philosophical and theoretical viewpoint (Baltes, Reese, & Nesselroade, 1977; Reese & Overton, 1970).

Four major philosophical ideologies (Romanticism, Cultural Transmission, Progressivism, and Holism) and, respectively, four associated developmental perspectives (maturational, behavioral, process-oriented, and ecological) have had an impact on early intervention practices (see Dunst, 1981, 1982b; Kohlberg & Mayer, 1972; Lambie, Bond, & Weikart, 1975; Stevens & King, 1976). Each ideology embraces a different definition of behavior change and development. Thus each ideology embraces a different way of defining what constitutes the most appropriate indices of progress. Table 1 summarizes the developmental perspectives and definitions of change and progress derived from each of these ideologies.

Maturational

This position views development as a progressive unfolding of behavior which reflects maturation of the central nervous system (CNS). Change in the functioning of the CNS is inferred from age-related but not necessarily interrelated developmental landmarks or behavioral responses. On a 10-item scale of reflex behavior, a child may on one occasion show only two reflex responses. If, on a second occasion, the child displays eight reflex behaviors, we may infer that a maturation of the CNS may have occurred. According to Gessell, graded series of behavior landmarks are the measuring rods or criterion of normal development. And, the manifestation of these behaviors in response to standardized test situations provides a basis for inferring the degree of maturity of the CNS. Similar concepts regarding changes in the CNS are espoused by Karl and Berta Bobath in their neurodevelopmental theory and Jean Ayres in her theory of sensory integration. Both perspectives view deve-

Table 1

Comparison of the Four Major Developmental Perspectives

Philosophical Ideology	Developmental Perspective	Definition of Development	Major Proponents
Romanticism	Maturational	Unfolding of innate predispositions resulting from maturation of the central nervous system.	Arnold Gessell, Catherine Amatruda, Carl and Berta Boboth, Jean Ayres
Cultural Transmission	Behavioral	Accumulation of learned behaviors as a result of reinforcement of the behavior.	B.F. Skinner, Albert Bandura, Sidney Bijou, Donald Baer, Todd Risley
Progressivism	Process-Oriented	Progressive changes in the manner in which information is acquired, stored, and used.	Jean Piaget, J. McVicker Hunt, Ina Uzgiris, Robert McCall, Kurt Fischer, Robby Case, Philip Zelazo
Holism	Ecological	Differentiation in topography of behavior and adaptations to environmental demands.	Urie Bronfenbrenner, Kurt Lewin, Moncrieff Cochran, Jane Brassard

lopment as the increasing capacity to master control of the CNS; maturity of the CNS is reflected in so-called normal patterns of movement.

Behavioral

According to this view, behavior is influenced by its environmental consequences; if behavior is followed by some reinforcing event, this strengthens the tendency to act or behave (Skinner, 1974). Behaviors which produce reinforcing consequences are called operants. The behavioral point of view considers behavior to be affected primarily by the environment.

Development from a behavioral perspective is defined in terms of the accumulation of learned behaviors: the more one learns, the more development has occurred. Change is most often measured in terms of how many behaviors have been learned in a given period of instruction.

Process-Oriented

This position views development as progressive changes in the manner or processes used to acquire, store, and use information. Changes in the particular types of intellectual operations used to solve problems or adapt to more complex demands are the principal indices used to measure developmental change. Piaget's theory of the progressive changes that occur in the genesis of intellectual capabilities is perhaps the most widely used process-oriented model for characterizing developmental change. According to Piaget (1970),

there are distinct levels or stages of development, each of which is characterized by progressively more sophisticated ways a child organizes, understands, and uses information to respond to the environment.

The information-processing theory (Zelazo, 1979) is another process-oriented concept of development and is beginning to influence early intervention practices. Development is considered to be a progression of changes in higher-order strategies for acquiring, storing, and using bits of knowledge and information.

Ecological

This position views development as a progressive differentiation of a person's conception and knowledge of environments and the capacity to discover, sustain, restructure, or alter environments at levels of similar or greater complexity (Bronfenbrenner, 1979). Development is considered to have occured if the form and content of a person's activities, roles, and interpersonal relationships become progressively more differentiated as a function of direct and indirect experiences. From an ecological perspective, progress and change are defined as differentiations and adaptations of behavior in response to demands and expectations of different persons and settings. According to Bronfenbrenner (1979), a shift in "balance of power" away from adults toward the developing child is one major indicator of behavior change. For example, the child's capacity to initiate and sustain interactions with others as opposed to relying on adults to prompt behavior is one measure of a shift in balance of power.

HOW CAN CHILD PROGRESS BE MEASURED?

A number of different strategies and approaches can be used to measure child progress and developmental change. Each is derived from one or more of the developmental views described above.

When we ask: "How can we measure child progress?" we really ask: "What should be the outcome measure of the intervention effort?" An outcome measure (dependent variable) is something we expect to change or affect as a result of some type of treatment (independent variable). Table 2 shows each of the major developmental views described in the previous section and provides examples of the methods of measurement and indices of behavior change associated with each view.

Norm-Referenced Scales

These standardized tests have been administered to large numbers of children for the purpose of determining the usual ages at which children reach different behavior landmarks. Standardized, norm-referenced instruments

Table 2

Methods Used to Measure Child Progress

Developmental Perspective	Measurement Technique	Focus of Assessment	Examples of Indices of Progress
Maturational	Norm-referenced scales	Comparisons of child against performance of standardization samples or standard established as expected outcome by the investigator.	Primary: Number of items passed, DA, MA, DQ, IQ, MDI Secondary: Gain scores; developmental rate scores; ratio of expected to actual progress
Behavioral	In vivo observation/measurement of behavior; behavior rating scales; criterion-referenced scales	Determination of baseline performance and change in performance as a result of intervention.	Changes in frequency, duration, rate, chaining of behavior; behavior engagement.
Process-Oriented	Ordinal scales; psychophysiological measurement of cognitive understanding; cognitive effectance and motivation	Determination of the changes in cognitive performance resulting from challenging experiences (intervention).	Changes in problem-solving abilities; shifts in cognitive capabilities; mastery of cognitive tasks; rates of progress.
Ecological	In vivo observation of behavior; social support rating scales; ratings of levels of independence; interaction ratings	Determination of direct and indirect influences that affect child performance.	Shift in balance of power; increased topography (differentiation) of behavior, adaptations to environmental demands, parental perception/expectations; situational outcomes (school placements, institutionalization, etc.)

include the Bayley Scales, Gesell's Developmental Schedules, the Cattell Scale, the Griffith's Mental Development Scales, the Stanford-Binet Scales, the McCarthy Scales, and the Vineland Social Maturity Scale.

Norm-referenced scales permit a comparison of a particular child's performance and the performance of other children included in the standardization sample. The norms established for a given assessment instrument indicate a "typical" performance and, thus, provide a basis for determining the extent to which a particular child's performance is advanced or delayed.

Primary indices. Norm-referenced tests yield three primary indices of performance: (1) total number of times passed, (2) developmental age(s), and (3) developmental quotients.

The total number of items passed on a test may be an appropriate measure of performance only if there are an equal number of items at each age level. If not, changes in the number of items passed on different measurement occasions will not reflect the same amount of progress for different children.

Total number of items passed may be an appropriate measure for scales like the Griffiths and Cattell tests; it is not an appropriate measure for tests like the Bayley scales.

Nearly all norm-referenced tests provide procedures for computing a developmental level of performance based on the total number of items passed by a child. This index provides a rough gauge for making statements like: "This child is functioning much like a child of 'X' chronological age." A child's developmental age (DA) and mental age (MA) are often used to determine how much progress or change has occurred between two separate measurement occasions. The child's DA or MA at a second measurement occasion is subtracted from that observed at a previous measurement occasion.

A third index that can be derived from norm-referenced tests is a development quotient (DQ).* This index is computed as a ratio quotient or a deviation quotient.

$$DQ = \frac{\text{Developmental Age}}{\text{Chronological Age}} \times 100$$

Deviation quotients are derived separately for different age groups of children in the standard sample. Tests that use deviation quotients always include tables for determining this index.

Ratio quotients have different meaning than do deviation quotients and generally should not be used with scales that use deviation quotients for gauging rate of development. Typically, scales that use deviation quotients do not provide quotients below certain cutoff points (e.g., 50 on the Bayley scales); thus making them less useful for low-functioning children. In such instances one is tempted to use the test manual tables (for computation of deviation quotients above the cutoff point) and the ratio method to determine deviation quotients below the cutoff point. This should not be done. The best solution is, on an a priori basis, to use a test that will accurately characterize DQs for children functioning at quite varied developmental levels. Alternately, one could compute ratio quotient scores for all subjects (if it is kept in mind that these indices are very rough measures of rate of development.) That is, if the children in a program perform both above and below the cut-off point for a given test, and one wants to have a common metric for contrasting rates of development, the child's mental age divided by his or her chronological age multipled by 100 will give you a general indication of each child's performance.

Derived indices. With regard to both the efficacy of intervention and measurement of child progress, this principal question is being asked: "Was child progress greater under conditions of intervention relative to a non-intervention condition?" Typically, we have some pretest scores and some posttest score on a group of children enrolled in an early intervention

---

* Here DQ is used interchangeably with IQ, MDI, GQ, and other indices of performance that provide a metric of the relationship between actual and typical performance.

program. For illustrative purposes, we will assume we have data on a child to whom a developmental test was administered at 12 and 18 months of age. The child's developmental ages at those measurement occasions were, respectively, 6 and 10 months. From birth to age one year, the child received no intervention services. From age 12 to 18 months, the child was enrolled in an early intervention program.

Computation of the child's ratio quotient for the two assessments shows that at age 12 and 18 months the child's developmental quotients were, respectively, 50 and 56. (This is a small and insignificant change if one considers the fact that the standard error of measurement for infant psychometric tests is generally 6 or more points.) Are we to conclude that intervention produced no differential progress between tests? The answer is yes -- if we use primary developmental indices (i.e., developmental ages) as the outcome measures. However, this conclusion is flawed; the 18-month developmental age reflects experiences during the intervention period and the lack of experiences during the nonintervention period. A better way to assess progress during the intervention period is to separate changes due to the intervention period and the nonintervention period. Use the following formula to discern differential progress:

$$\frac{(DA_2 - DA_1)}{(CA_2 - CA_1)} \quad / \quad \frac{(DA_1)}{(CA_1)}$$

DA_1 and DA_2 are the child's developmental ages at the first and second measurement occasions, respectively. CA_1 and CA_2 are the child's chronological ages at the respective measurement occasions. For our example,

$$\frac{10 - 6}{18 - 12} \quad / \quad \frac{6}{12} \quad = \quad .67 \; / \; .50 \quad = 1.34 \text{ gain}$$

Thus, this particular child made more progress during the intervention period than the nonintervention period, as shown by the ratio of expected (denominator term) to observed (numerator term) change.* We can conclude that something during the intervention period affected the child's development.

The above formula is a derivative of the ratio deviation method for computing rate of development. This formula, however, separates out changes for the intervention and nonintervention periods. Similar indices of progress are described by Bagnato and Neisworth (1980) and Simeonsson and Wiegerink (1975).

Despite advantages over primary indices, the above derived index has one shortcoming. The computational methods are based on the assumption that normal progress constitutes a unit increase in DA for every unit increase in CA (i.e., development is linear), and that acceleration of development equal to "normal" can be a reasonable goal for handicapped children. That is, a one-

---

* A ratio score exceeding 1.0 shows that more than expected progress was made whereas a score less than unity indicates that amount of progress was less than during the nonintervention period.

month gain for each month in the program is expected under this assumption of normal development. Under most conditions this is a tenable and useful assumption. However, with severely handicapped youngsters, this assumption is likely to result in a conclusion that significant change was not made when, in fact, it was. An alternative is to use predicted progress based on the pretest scores of a group of handicapped youngsters (see especially Bricker, Sheehan, & Littman, 1981), rather than "normal" development, as the benchmark. Scores obtained by a group of children at entry into a program are used to predict scores at a later time. The discrepancy between the predicted and actual scores provide a basis for gauging program impact. The procedure involves the regression of DA on CA for pretest scores (scores obtained prior to intervention), the derivation of a regression equation to predict DA from CA, the computation of a predicted DA from the CA when posttest was administered, and a comparison of the actual DA with the predicted DA at this CA.

Most early intervention programs use standardized tests as one index of child progress. The above methods are approaches for computing progress from these scales. There are, however, other indices of development that might be used as measures of child progress.

Process-Oriented

Ordinal scales. Ordinal, or Piagetian-based scales, can be used to measure developmental change. These scales include series of items ranked by order of difficulty. Successively higher levels of achievement represent progressively more complex forms of behavior in the genesis of different types of early cognitive competencies. For example, the Uzgiris-Hunt (1975) scales measure performance in seven separate sensorimotor areas: object permanence, means-ends abilities, vocal imitation, gestural imitation, causality, spatial relationships, and schemes for relating to objects (play). Developmental status on ordinal scales is determined by noting the highest item passed (i.e., landmark achieved) on each separate sensorimotor scale.

Because there is considerable specificity in early sensorimotor development (Hunt, 1976), progress and rate of development can vary between different sensorimotor domains. Thus, if one wanted to assess the efficacy of vocal imitation training, progress along that particular scale could be compared to progress in other sensorimotor areas and differential effects of the training efforts could be determined.

Dunst and Gallagher (1983) recently described a strategy to use ordinal scales that has particular import for assessing the efficacy of early intervention efforts. The average ages at which successive ordinal landmarks are achieved for an intervention population are assessed. These mean ages and ages of infants in other samples for whom comparable data have already been collected are compared (see e.g., Hunt, 1976). Dunst (Dunst & Gallagher, 1983) used this approach to show that a biologically impaired group of infants in a home-based early intervention program reached the top-level landmarks on an object permanence scale before biologically impaired and nonimpaired institutionalized subjects. This suggested that intervention affected, in part, the patterns of development of the children. Other types of comparisons can

also be made (e.g., age of acquisition of landmarks for infants entering intervention programs early vs. late: different types of intervention; and different levels of intervention).

Physiological measures. The use of heart rate as a process-oriented outcome measure is worthy of special note since it offers promise of being an index of change with children for whom psychometric, ordinal, and other overt behavior indices are inappropriate and discriminatory. Zelazo and his colleagues (see e.g., Kearsley, 1979; Zelazo, 1979) demonstrated that heart rate, when used as a measure of attention (together with visual fixation), often shows that developmentally disabled children are not as delayed in cognitive understanding as the results of traditional approaches to assessment seem to indicate. The following paradigm is described by Zelazo:

> The primary objective is to present an engaging visual or auditory event and to repeat that event until the child creates an expectancy for it. After a fixed number of presentations, and generally before the child has lost all interest in the event, a moderately discrepant variation of the standard is introduced and too is repeated. Following three presentations of the discrepancy, the standard is re-introduced for three presentations. (pp. 239-240)

The child's capacity for creating an expectancy for the standard assessment, whether or not the child recognized the reappearance of the standard following the discrepancy, and whether or not the child can assimilate the discrepant variation itself, are measured.

Initial presentations of an event cause heart rate to decrease (signalling attention) but stabilize across repeated presentations of the same stimulus. Presentation of the discrepant event causes heart rate to decrease again if the child recognizes the events as being different. Zelazo has collected cross-sectional and longitudinal data on children between age 3 and 36 months which provide a basis for deciding "level of cognitive understanding." Matching the responses of a disabled child to those of nondisabled children provides a basis for deciding the child's information processing capabilities. At present, this approach to assessment has been used only to discern level of cognitive integrity for determining what constitutes an appropriate "entry level" for intervention. Whether or not it might be used to assess child progress is open to investigation -- but does seem that it has potential utility for this purpose, if it could be shown that following intervention the child's information processing capabilities showed advances relative to baseline performance.

Behavior Measures

Use of psychometric tests and, in some cases, ordinal scales assumes that an intervention is designed to affect the acquisition of the types of competencies that these scales measure. Often, this is not a good assumption. For example, for a program that is designed to affect changes in the communicative abilities of young children, the Bayley scales, Uzgiris-Hunt scales, or any other test of infant intelligence would be an inappropriate outcome

measure. In such cases, measures more specific to the goal of the intervention would be better indices of behavior change.

The following five behavior indices can be used as outcome measures when tailored to specific interventions:

- Frequency -- number of occurrences of a specific behavior during a specified period of time (e.g., number of tantrums per hour)
- Duration -- absolute amount of time spend engaged in a specified activity (e.g., amount of time engaged in cooperative play)
- Proportion -- differential amount of time spent in each of two or more activities (e.g., time spent in tantrums vs. time spent in cooperative play)
- Rate -- rate of change from one activity to another (e.g., number of changes from tantrums to cooperative play to tantrums)
- Sequence -- pattern of occurrence of two or more behaviors (e.g., tantrums, then cooperative play, then pushing, then crying, then tantrums, etc.)

"Behavior engagement" (e.g., Bailey, Harms, & Clifford, 1983; Montes & Risley, 1975) is a measure of the efficacy of early intervention which has received increased attention and which deserves special comment because of its general usefulness. Engagement refers to the amount of time a child plays with materials, interacts with peers and adults, and otherwise remains actively involved with his or her environment in a developmentally appropriate manner (Dunst, in press). This duration measure of behavior differs from the typical frequency of production of behavior that has generally been used as the measure of child progress (e.g., Shearer & Shearer, 1972). To the extent that a child is actively engaged with his or her environment, opportunities abound for the possibility that the child will acquire increased variation in his or her topography of behavior (role differentiation) and will learn to adapt to the demands of different settings, persons, contexts, etc. (Bronfenbrenner, 1979). Engagement would appear to be a particularly important measure of the efficacy of early intervention efforts if it could be demonstrated that engagement levels varied as a function of intervention efforts.

Ecological-Based Outcome Measures

An ecological perspective of development, because of its orientation, suggests a number of additional outcome measures for assessing child progress. As noted by Bronfenbrenner (1979), early intervention programs have typically been evaluated using ecologically constricted outcome measures. This means that we have used outcome measures that provide little basis for knowing about the child as he or she functions in situations of everyday life. Moreover, Bronfenbrenner (1979) notes that we have almost exclusively focused on the child as the experimental subject with little or no attention paid to the interaction of child and others -- both animate and inanimate -- in assessing progress and behavior change.

Shift in balance of power. According to Bronfenbrenner (1979), one index of development is the shift of power during child-adult interactions away from the adult toward the developing child. As a child acquires more socially

adaptive behaviors and becomes progressively better able to initiate and sustain interactions with others, one would expect the child to assert more control over his or her environment, with a concomitant decrease in the need for an adult to do so for the child. This decisively simple contention suggests a relatively powerful means for assessing child progress. A measurement scale for determining developmental changes could be something as simple as a list of "initiating behaviors" ranked by degree of difficulty with a four- or five-point rating scale that allows one to assess the degree to which the behaviors are exhibited by the child and others within and across settings. Obtaining these measures across several months would provide a basis for determining the amount and type of shifts that occur in the balance of power between the child and others.

Role differentiation. This outcome measure reflects the fact that development involves increased variation in the types of behaviors acquired as a result of experience and the ability to adapt to the demands of different settings and contexts by either using what one already knows or constructing new behavior. Intervention efforts could define the types of differentiated roles expected to be acquired as a function of the intervention. (Construct a simple recording system that collects data on predicted and nonpredicted behavior outcomes, and compare the findings to see if expected changes exceeded changes for nonpredicted behaviors.) By tracking the topography and situation-specific manifestation of behaviors over time, the rate of occurrence of behaviors can be plotted across situations to assess the net increase in behaviors that are learned as a result of involvement in an intervention program.

Changes in perceptions of behaviors. According to Bronfenbrenner (1979), when developmental change occurs, it takes place concurrently in two domains -- perceptive and active. I would argue strongly that some type of intervention may affect perceptions of a child's behavior. This, in turn, may affect actual behavior manifested toward the child. And this, in turn, is likely to affect child behavior. A study conducted in the Child Development Laboratory at Western Carolina Center revealed that parents with large degrees of social support available to them perceived their children's handicaps as being less devastating than did parents who did not have a lot of support (Dunst, Trivette, & Cross, 1984). Changes in perceptions may be a "first order" effect of intervention efforts, and "second order" effects include actual child change.

Situational outcomes. Lazar and Darlington (1982) demonstrated that situational and contextual outcome measures may be good indicators of the success of intervention efforts. For example, environmentally at-risk children enrolled in early intervention projects were found to be less likely to be placed in special education classes or be retained in grade than were at-risk children who were not enrolled in early intervention projects. These types of outcomes suggest a number of ecologically relevant measures of the efficacy of early intervention. For example, Deno's (1970) cascade of services model suggests an ordinal scale of possible placement options (regular class, special education class, home-bound, etc.) that could be used as an outcome measure of the effects of early intervention services. At Western Carolina Center, for instance, we have found that a handicapped child who did not receive early intervention services is twice as likely to be institu-

tionalized as a child who did receive early intervention services. The use of "placement outcomes" as measures of program impact would provide considerable ecological validity to our evaluative efforts, and such measures are strongly indicated as indices of efficacy.

WHAT ARE SOME STRATEGIES FOR DOCUMENTING THE EFFECTS OF EARLY INTERVENTION?

The basic question being asked when evaluating the efficacy of intervention is: "Did the intervention (independent variable) have the effect (dependent variable) expected?" There are two broad types of efficacy studies: those that assess the impact of a specific intervention effort (e.g., a cup drinking training program) and those that assess the overall or general impact of an early intervention program. Strategies designed to assess the general effects of intervention efforts are described here.

This section addresses three aspects of evaluative studies: defining the independent variable, asking the appropriate questions, and research design issues to answer evaluation questions.

Defining the Independent Variable

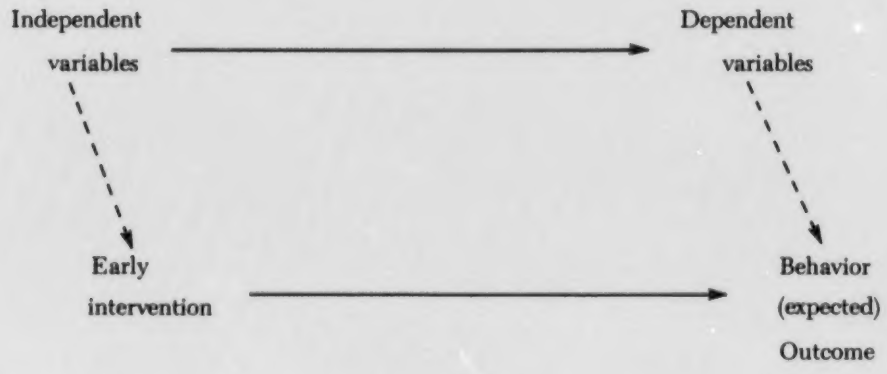
The independent variable is the thing that is done that is expected to produce some change or outcome (dependent measures). Figure 1 (next page) shows the relationship between early intervention as an independent variable and the outcomes (dependent variables) that might be expected to be realized. The independent variable is the intervention which is hypothesized to produce the types of changes described in the previous sections of this chapter. All of the indices described as measures of child progress are the types of outcomes that one might hypothesize to be affected by the intervention, though good evaluation requires one to specify the exact types of changes anticipated as a result of a particular intervention.

The preferred or recommended approach to establish the efficacy of a treatment or intervention is to compare children who received early intervention and those who did not. This requires an experimental design with a control group where subjects are randomly assigned to intervention and nonintervention conditions. However, the typical approach that has been taken is to use a pretest-posttest design where an intervention population is tested at Time 1 and then again at Time 2, and the difference between the measurement occasions is assumed attributable to the intervention (see Dunst & Rheingrover, 1981). There are a number of assumptions of both designs that are not generally tenable, and thus make their use questionable for answering the efficacy question (see especially Dunst, in press).

First, there is the assumption that all individuals who constitute the intervention populations benefit equally from the intervention. Second, there is the assumption that the amount of intervention provided the intervention population is similar for all individuals. Third, there is an assumption that

Figure 1

Early Intervention and Expected Outcomes as Independent and Dependent Variables



defining the independent variables as "'involvement' in an early intervention program" accurately describes the events that affect child outcomes.

I would argue that these assumptions are often untenable and, as a result, decrease the likelihood that benefits can be demonstrated even when such benefits have in fact occurred. The above assumptions are factors that contribute to increased variability among subjects within groups, and it is this increased variability that makes it more difficult to statistically demonstrate the benefits of early intervention. As it turns out, we can take advantage of this variability in assessing the impact of early intervention. First, note the fallacies underlying the above three assumptions.

The assumption about the homogeneous effects of early intervention. This assumption supposes that most children benefit equally from early intervention. Efforts to document the efficacy of early intervention with biologically impaired infants almost never take into consideration the probability that the intervention will have different impacts depending upon the levels of functioning, degree of retardation, or the handicapping conditions of the subjects. Thus, an outcome measure would be expected to show substantial variability; this, in turn, decreases the probability of finding statistically significant results.

The assumption about the intensity of intervention. This assumption supposes that the degree of involvement and intensity of intervention is the same for most program participants. This assumption is implicitly made

any time the intervention program, rather than specific dimensions of the program (e.g., classroom attendance, number of home visits), is used as the basis for defining the independent variable. Intensity of involvement varies considerably for a host of reasons, including parental level of interest, the child's propensity to bouts of illness, transportation problems, various family conditions, the child's degree of impairment, etc. Since intensity has been found to be a predictor of progress (Sandow & Clarke, 1978), making the assumption about similarities in level of involvement seems untenable.

The assumption about early intervention as an independent variable. This assumption supposes that the early intervention program is the principal or only intervention provided to the program participants. This assumption seems particularly untenable for a number of reasons. First, parents who seek out early intervention services are likely to access other services for their children (e.g., health-related activities) and themselves (e.g., parent support groups). Second, social system theory strongly suggests that early intervention is only one of a number of support services (both informal and formal) that are likely to have some impact on the child (Cochran & Brassard, 1979) as well as family functioning (Gabel, McDowell, & Cerreto, 1983). A number of implications are made when conceptualizing early intervention as the principal or only treatment provided to children and their families. Methodologically, we may mistakenly attribute changes on the dependent measures to the early intervention program when, in fact, changes were related to other support services or an interaction between the program services and other types of support. In fact, many programs actively encourage other support (both formal and informal) to children and their families in addition to the services provided by the program. The failure to take this into consideration as part of defining "early intervention" as an independent variable is likely to reduce the probability of demonstrating program effectiveness. A better strategy for discerning the impact of intervention is to consider early intervention as one of a series of possible support services, to consider the various support services as an aggregate of "early intervention," and then to determine the relative contributions of this composite independent variable on the dependent measures of interest. This approach, described in more detail below, suggests a methodologically strong and ecologically realistic perspective of the real-world forms of support which have an impact on the child and family.

Alternative Ways of Conceptualizing the Independent Variable

The problem with the assumption that interventions have no differential effects can be addressed by dividing the intervention population into subgroups that decrease the variability within groups and thus increase the possibility of at least demonstrating differential effects. Groups could be constituted on such variables as diagnosis, levels of mental retardation, early vs. late entry into the program, or any other variable that divides the group into subgroups for the purpose of assessing how group membership is related to some desired outcome.

The problem with the assumption that all individuals receive the same amount of intervention can be addressed by dividing the intervention population into subgroups according to degree of involvement in the program. For

example, in a center-based program, children could be grouped by days per week attending the program. Likewise, in a home-based program, number of home visits conducted during the year might be used to divide the group.

The problem with the assumption that the intervention is the only or primary event affecting an outcome measure can be addressed by treating the intervention as only one of several independent variables that are related to the dependent variables of interest. For example, a simple rating scale could be developed that measures the degree to which a child and his or her family receive a host of services from the intervention project and other agencies. By constructing the scale in a way that the children and families have equal access to the various services, either the total of the rating scores or the individual scores for separate services could be used as independent measures.

An approach we have taken at the Family, Infant and Preschool Program is to have the parents themselves define their level of involvement and satisfaction with various sources of support. This is used as the independent variable. The scale consists of 18 potential sources of support to the families (Dunst, Jenkins, & Trivette, 1984). The parents rate, on a five-point scale, the degree to which each source has helped them care for and rear their handicapped child. The total of the 18 separate ratings and the ratings for the individual items are used as independent measures.

There are a number of advantages for taking this approach to program evaluation. First, the approach is based on the sound assumption that parental ratings of the helpfulness and benefits derived from various sources is one way of defining the independent variable. Our own research supports this assertion. In several studies we have conducted, amount of support available to families of handicapped youngsters has been found to be significantly related to a number of parental, family, and child outcome measures. Second, if different programs use the same scale, the programs could combine data across projects to build a sufficient sample to perform a variety of statistical tests which otherwise could not be performed -- a problem that has plagued early intervention efficacy studies (see Dunst & Rheingrover, 1981).

Posing Appropriate Questions

The reader should note that the above strategies do not ask the question: "Does early intervention work?" Rather, the strategies are designed to answer the question: "What particular aspects of interventions and other forms of support affect program outcomes?" Note also that the latter question is designed to address issues regarding the specificity of intervention effects and the unique contributions different variables make to different outcomes. Evaluative efforts that answer this question provide specific information regarding what does and does not affect program outcomes. This information can demonstrate the efficacy of intervention efforts and can be useful when justifying improvements and changes in program operations.

Posing the proper question in terms of the relationship between independent and dependent variables is also an important part of evaluative studies. Generally, evaluation research in the field of early intervention has been

conducted in an atheoretical manner. This is the case because the relationship between what one does (independent variable) and what one expects to occur (dependent variable) has not been well conceptualized. For the most part, the field of early intervention has not adequately defined what constitutes "early intervention." Further, and as a result of a failure to define the independent variable, outcome measures often assess competencies that may or may not be assumed to be affected by what one does under the rubric of early intervention. For example, most early intervention programs use standardized instruments like the Bayley scales or nonstandardized instruments like the Early LAP (see e.g., Garwood, 1982). If so many projects are designed to develop innovative efforts, why do most of these projects use the same or similar outcome measures!? Better conceptualized studies are sorely needed if we are to begin to unravel what it is that affects different program outcomes.

Several Generally Feasible Research Strategies

The types of questions just posed do not require nontreatment control groups to evaluate the effects of intervention. Because the key question asked is "What aspects of early intervention affect program outcomes?" no control groups are needed and yet the influence of specific variables can be determined. Two generally feasible approaches for doing so are described below.

Split-group design. In those instances where an intervention population can be divided into two or three subgroups on the basis of some important explainer variable, a comparison can be made of the differential effects of intensity of treatment, level of involvement, parental satisfaction, etc. For example, if level of intensity were predicted to affect program outcomes, and the participants in the program varied according to the amount and time of services they received, a median split of the group into low- and high-intensity subgroups could be made. Then the two subgroups could be compared to determine if in fact the high-intensity group made significantly greater gains, changes, etc. If this were confirmed, we would then have evidence to suggest that level of intensity was a significant contributor to change on the dependent measures.

The split-group design has general utility as long as the method of grouping subjects on the independent variable is based entirely on the type of predicted relationship between the independent and dependent measures. The split-group design can be used whenever one can conceptually predict such relationships. For example, at the Western Carolina Center, an ecological-social support model is used to structure intervention efforts. The model predicts that social support available to families affects parental well-being, family integrity, styles of interaction between parents and children, and selected aspects of child behavior. Parents complete a simple social-support rating scale which is used to group the parents according to low, median, or high degrees of support. Group membership is then related to the above classes of dependent measures (Dunst, 1982a). This strategy has allowed us to identify a number of different and revealing relationships between social support and the outcomes which are expected to be influenced by support.

Multiple regression analysis. In those instances where we have a series of independent measures and we want to determine the unique contributions of each in terms of affect (for example, child outcomes) a multiple regression analysis can help (see especially Cohen & Cohen, 1983). Multiple regression is a general method for assessing the relative contributions of a series of different independent measures to one dependent meaure. Two approaches to regression analysis seem to have utility for the field of early intervention.

Method 1. Assume that we have a relatively distinct set of variables that we predict have some affect on a specific child outcome. The predictor or independent variables might be child sex, number of individual activities prescribed, precentage of home visits made, ratings of parental commitment to the interventions, etc. The strategy to determine the relationships between each predictor variable and the child outcome measure involves performing a multiple regression analysis where child outcomes (Y) is predicted from the optimal linear estimates of combinations of the independent variables (X_1, X_2, X_3, X_n). The analysis can proceed in one of three ways. First, a full regression model can be used where all the independent variables are simultaneously regressed on the dependent measure. A significant result would indicate that the combination of predictors affect the outcome in a manner which suggests a casual or mediational relationship between the independent and dependent variables. Second, a stepwise regression analysis can be performed. In this approach, the independent variables are entered one at a time until no independent variable accounts for a significant amount of variance in the dependent variables. The order of entry provides a basis for discerning the relative contribution of the independent variables. The independent variable most related to the dependent variable is entered first, the next most related independent variable is entered second, and so on. Third, a hierarchical regression analysis can be used. In this method, the order of entry of independent variables into the equation is determined on an a priori basis and dictated by the nature of the evaluation effort. At each step in the analysis, the increments (I) in R^2 are determined and the significance of the variance accounted by I is determined. Thus, the extent to which specific independent variables are related to the dependent variable -- and the specific independent variables that are important predictor variables in terms of affecting changes in the dependent variable -- can be assessed.

Method 2. The second method for assessing the contributions of a series of independent variables is to use multiple regression analysis where the independent measures are conceptually or programmatically similar. For example, the independent meausures might be ratings of the degree to which parents find, for example, that each of 10 sources of support helps them care for their child. Or, staff or parents might rate the extent to which children are involved in a number of different services offered by a program. The simultaneous, stepwise, or hierarchical models can be used to discern the relationship between the independent variables and the dependent variable, although only the hierarchical approach can tell which specific independent variables are most important. Thus, the hierarchical approach is recommended. The reader is referred to Dunst (in press) and Dunst & Trivette (1984) for more in-depth discussions and illustrations of this approach to program evaluation.

CONCLUSION

The extent to which the efficacy of early interention can be established in terms of its impact upon child progress depends upon how well the relationship between the intervention (independent variable) and outcome measures (dependent variable) are specified. Especially crucial for good program evaluation are adequate definitions of the independent variables, recognition of the assumptions inherent in one's conceptualization of the intervention, a selection of outcome measures that one would expect to be affected by the intervention, and methods for discerning the impact of the interventions.

It is hoped that the various discussions presented here stimulate the reader to better design program evaluation studies. By doing so, the effects of early intervention on child progress can be accurately discerned.

REFERENCES

Bagnato, S., & Neisworth, J. (1980). The Intervention Efficiency Index: An Approach to Preschool Program Accountability. Exceptional Children, 46, 264-269.

Bailey, D., Harms, T., & Clifford, R. (1983). Matching Changes in Preschool Environments to Desired Changes in Child Behavior. Journal of the Division for Early Childhood, 7, 61-68.

Baltes, P., Reese, H., & Nesselroade, J. (1977). Life-Span Developmental Psychology: Introduction to Research Methods. Monterey, California: Brooks/Cole.

Bricker, D., Sheehan, R., & Littman, D. (1981). Early Intervention: A Plan for Evaluating Program Impact. Monmouth, Oregon: Western States Technical Assistance Resource.

Bronfenbrenner, U. (1979). The Ecology of Human Development: Experiments by Nature and Design. Cambridge, Massachusetts: Harvard University Press.

Cochran, M., & Brassard, J. (1979). Child Development and Personal Social Networks. Child Development, 50, 601-616.

Cohen, J., & Cohen, P. (1983). Applied Multiple Regression Correlation Analysis for the Behavioral Sciences (2nd edition). Hillsdale, New Jersey: Erlbaum..

Deno, E. (1970). Special Education and Developmental Capital. Exceptional Children, 39, 229-237.

Dunst, C.J. (1981). Infant Learning. Hingham, Massachusetts: Teaching Resources.

Dunst, C.J. (1982a). Social Support, Early Intervention, and Institutional Avoidance. Paper presented at the annual meeting of the Southeastern Association on Mental Deficiency, Louisville, Kentucky.

Dunst, C.J. (1982b). Theoretical Bases and Pragmatic Considerations in Infant Curriculum Construction. In J. Anderson (Ed.), Curricula for High-Risk and Handicapped Infants. Chapel Hill, North Carolina: Technical Assistance Developmental System (TADS).

Dunst, C.J. (in press). Overview of the Efficacy of Early Intervention Programs: Methodological and Conceptual Considerations. In L. Bickman & D. Weatherford (Eds.), Evaluating Early Intervention Programs for Severely Handicapped Children and their Families. Austin, Texas: PRO-ED.

Dunst, C.J., & Gallagher, J. (1983) Piagetian Approaches to Infant Assessment. Topics in Early Childhood Special Education, 3(1), 44-62.

Dunst, C. & Rheingrover, R. (1981). An Analysis of the Efficacy of Infant Intervention Programs with Organically Handicapped Children. Evaluation and Program Planning, 4, 287-323.

Dunst, C.J., Jenkins, V., & Trivette, C.M. (1984). Family Support Scale: Reliability and Validity. Journal of Individual, Family, and Community Wellness, 1(4), 45-52.

Dunst, C.J., & Trivette, C.M. (1984). Differential Influences of Social Support on Mentally Retarded Children and their Families. Paper presented at the annual meeting of the American Psychological Association, Toronto, Ontario, Canada.

Dunst, C.J., Trivette, C.M., & Cross, A. (1984). Mediating Influences of Social Support: Personal, Family, and Child Outcomes. Manuscript submitted for publication.

Gable, H., McDowell, J., & Cerreto, M. (1983). Family Adaptation to the Handicapped Infant. In S.G. Garwood & R. Fewell (Eds.), Educating Handicapped Infants. Rockville, Maryland: Aspen Systems Corporation.

Garwood, S.G. (1982). (Mis)use of Developmental Scales in Program Evaluation. Topics in Early Childhood Special Education, 1(4), 61-69.

Hunt, J. McV. (1976). Utility of Ordinal Scales Derived from Piaget's Observations. Merrill-Palmer Quarterly, 22, 31-45.

Kearsley, R. (1979). Iatrogenic Retardation: A Syndrome of Learned Incompetence. In R. Kearsley & I. Sigel (Eds.). Infants At Risk: Assessment of Cognitive Functioning. Hillsdale, New Jersey: Erlbaum.

Kohlberg, L., & Mayer, R. (1972). Development as the Aim of Education. Harvard Educational Review, 42, 449-496.

Lambie, D., Bond, J., & Weikart, D. (1975). Framework for Infant Education. In B. Friedlander, G. Sterritt, & C. Kirk (Eds.), Exceptional Infant (Vol. 3): Assessment and Intervention. New York City: Brunner/Mazel.

Lazar, I., & Darlington, R. (1982). Lasting Effects of Early Education. Monographs of the Society for Research in Child Development, 47 (No. 2-3), Serial No. 195.

Montes, F., & Risley, T.R. (1975). Evaluating Traditional Day Care Practices: An Empirical Approach. Child Care Quarterly, 4, 208-215.

Piaget, J. (1970). Piaget's Theory. In P. Mussen (Ed.), Carmichael's Manual of Child Psychology (3rd ed.). New York City: Wiley.

Reese, H., & Overton, W. (1970). Models of Development and Theories of Development. In L. Goulet & P. Baltes (Eds.), Life-Span Developmental Psychology: Research and Commentary. New York City: Academic Press.

Sandow, S., & Clarke, A.D. (1978). Home Intervention with Parents of Severely Subnormal, Preschool Children: An Interim Report. Child: Care, Health and Development, 4, 29-39.

Shearer, M., & Shearer, D. (1972). The Portage Project: A Model for Early Childhood Education. Exceptional Children, 36, 210-217.

Simeonsson, R., & Wiegerink, R. (1975). Accountability: A Dilemma in Infant Intervention. Exceptional Children, 41, 474-481.

Skinner, B.F. (1974). About Behaviorism. New York City: Knopf.

Stevens, J., & King, E. (1976). Administering Early Childhood Programs. Boston: Little, Brown and Co.

Uzgiris, I., & Hunt, J. McV. (1975). Assessment in Infancy: Ordinal Scales Psychological Development. Urbana: University of Illinois Press.

Zelazo, P. (1979). Reactivity to Perceptual-Cognitive Events: Application for Infant Assessment. In R. Kearsley & I. Sigel (Eds.), Infants at Risk: Assessment of Cognitive Functioning. Hillsdale, New Jersey: Erlbaum.

Chapter 3
Sources of Difficulty

by Nancy Johnson-Martin

WHAT IS CHILD PROGRESS?

Most interventionists probably agree that the measurement of child progress is critical both for planning appropriate interventions for individual children over time and for assessing the effectiveness of the programs that provide intervention services. Little agreement exists, however, when deciding what to include in measures of child progress for these two purposes. In defining child progress within this context, at least three sources of difficulty exist: 1) the discrepancy between implicit and explicit goals in programs serving handicapped infants; 2) the diversity of the needs of children served in early intervention programs; and 3) the different kinds of training of professionals who staff early intervention programs.

Implicit vs. Explicit Goals

Most early intervention programs define developmental progress in terms of the number of developmental milestones achieved by children over a given period of time. Thus, the explicit statement of goals primarily involves increasing the rate of developmental progress of these youngsters. A visit to any intervention program, however, would convince an observer that skill mastery is not the sole focus of early intervention. Parents are consoled, counseled, and trained; time is taken simply to entertain or play with the children; adaptive equipment is developed and modified. Children sometimes appear to be ignored as parents and interventionists discuss the meaning of certain responses, ways to deal with baby sitters, how to qualify for food stamps, etc. The time spent in these activities is not wasted. Rather, this time is used by staff and parents to meet implicit goals for themselves and for the children.

To assess all aspects of child progress related to an intervention program, one must consider the way time is spent within the program and the possible effects on the children. For example, time spent training and counseling parents is often directed at helping the parents feel competent to deal with their children and to interact with them in a reciprocal fashion. Though the focus is on the parents, the children might be expected to change as a result of interactions with the parents.

The definition of child progress should be broad enough to include the changes in children that occur as a result of better parenting or as a result of a generally less stressful environment. Items should be included in progress measures that reflect how the child is evolving as a member of his or her family and community. Such a view presents difficult measurement problems but, in the long run, will be more effective than methods now used to describe the progress of children in early intervention programs.

Needs of Children

Some of the difficulty in assessing progress in early intervention programs stems from the history of the early intervention movement. The first targets of intervention were children who were considered to be deprived of adequate stimulation and who could, therefore, benefit from an education or conpensatory stimulation program. Some of these children were born into impoverished homes, and some spent their early months in sterile hospital environments because of prematurity or other complications of pregnancy or delivery. For these children, a stimulation model was appropriate and child progress could reasonably be described by how clearly the child's development paralleled that of his or her nondeprived peers. When children with clearly atypical patterns of development became the targets of early intervention, however, this strategy often proved inadequate. These children needed more than stimulation; they needed active treatment or therapy.

In some instances, the effect of intervention might not be the acceleration of development; it may be the prevention of secondary handicapping conditions. For some of these children, "normal" interaction with the environment will never be possible. The focus of intervention may be to teach them atypical but functional ways to interact with the world (e.g., using manual signs or a communication board instead of, or in addition to, speech).

Unfortunately, the measurement techniques found useful with children at risk for developmental problems because of environmental factors were adopted somewhat uncritically by programs serving children with handicaps due to biological factors. The Intelligence Quotient (IO) or Mental Development Index (MDI) fixation has plagued early intervention programs by focusing attention on measures that may be reliable and scientifically respectable for some children, but insensitive to the progress made by atypical children.

Professional Training

In the beginning of the early intervention movement for biologically impaired children, programs tended to have an educational or a medical orientation. The basic goal was to promote developmental change, and the training and philosophical stance of the service providers, rather than the nature of the children, were often considered to be more important determinants of the way this goals was defined and implemented. Programs with an educational orientation focused attention on learning and behavior. Developmental tasks were analyzed and each step was taught with appropriate reinforcement. Programs with a medical orientation were more concerned with underlying

neurological development, the quality of movement patterns, and the concept of "readiness" for mastering tasks.

Most intervention programs today adopt a much greater integration of these two points of view, though some of the problems in defining progress stem directly from the philosophical and treatment issues that separate therapists from behaviorists. These issues are rarely divisive for children who are mildly to moderately impaired and whose development will follow typical patterns (though, perhaps, at a slower rate). Problems surface when the children are markedly atypical. The behaviorist's tendency to describe progress in terms of steps toward mastery of a specific task may be antithetical to the therapist's effort to build on underlying skills. For example, when teaching a child to sit alone, the behaviorist may attempt to place the child in a sitting position, propping as much as necessary, then gradually removing the props and measuring progress by how long the child is able to remain upright. The therapist, however, may not place the child in a sitting position at all. Rather, the therapist may work on a variety of activities in the prone position to strengthen trunk muscles that will later allow the child to sit unsupported. Progress toward sitting may be measured by how well the child masters the skills necessary to reach a sitting position unassisted.

When working with the most severely handicapped children, the goals of the therapist and the behaviorist may differ even more sharply. The behaviorist will tend to set goals that indicate skill mastery. The therapist may set goals that are primarily related to avoiding muscle contractures, maintaining range of motion, maintaining body functions, etc. From the therapist's standpoint, "not getting worse" can be considered progress -- a concept difficult for the behaviorist to accept and justify.

Before program staff can select instruments or techniques for measuring child progress, they must first examine their program and reach some thoughtful agreement about its goals and objectives. The implicit goals (goals implied by the way time is actually spent) should be made as explicit as possible. The next step is considering how children in the program might be affected by the goals that focus directly on them and by the goals that focus on their families. The program staff will then be ready to develop a definition of child progress that is broad enough to reflect both the unique characteristics of the children served and the activities that are a part of the intervention program. With this definition in mind, measurement instruments and techniques can be selected that are appropriate for assessing child progress within the context of that program.

HOW CAN CHILD PROGRESS BE MEASURED?

As noted earlier, the primary purposes for measuring child progress are to track the development of individual children so appropriate interventions may be planned and to assess program effectiveness. Each of these purposes requires appropriate measurement instruments.

Tracking the Progress of Individual Children

When the purpose of assessment is tracking a child's progress to plan appropriate interventions, the measurement instruments chosen should not be biased against the child's particular handicaps; the instruments should also reflect the desired content of intervention. Several instruments have been designed specifically for children who are sensorily impaired and cannot be assessed with standard instruments (e.g., the Reynell-Zinkin Scales--1979--for Young Visually Handicapped Children). Although such tests generally have poor norms, they are appropriate for tracking the progress of individual children because they are based on a knowledge of the "normal" development of atypical populations. For children with severe motor impairments, no such instruments exist. There are, however, a few criterion-referenced instruments that may be modified for use with children having various physical limitations (e.g., the Assessment Log of the Carolina Curriculum for Handicapped Infants--Johnson-Martin, Jens, & Attermeier, in press).

Children with multiple handicaps present special assessment problems. The clinician often must choose items from a variety of instruments and modify them to accomodate the child's particular handicapping condition. This type of modification invalidates the test norms; however, if the goal is simply to chart child progress, such invalidation is not critical.

Assessing Program Effectiveness

Assessment for the purpose of documenting program impact on child development involves much more than simply charting child progress. For almost all children, some developmental progress will be made regardless of whether any intervention has been provided. When documenting progress that results from intervention, total progress must be separated into two components: 1) progress due to development and 2) additional progress due to interaction between the intervention and the underlying developmental process.

When instruments are chosen to document intervention-related progress, many factors should be kept in mind. For example, what research design will be used to document intervention effectiveness? Will the design be one in which two groups of children are compared, one in which the progress of a group of children is compared against some standard, or one in which intervention effectiveness is documented in individual children and them summarized for the group (a single-subject design)? Additionally, if groups of children are to be compared, what are the characteristics of the children in each group? Are they relatively homogeneous (e.g., all have Down's syndrome) or heterogeneous? Is any one instrument appropriate for all the children? What instruments (if using a single-subject design) will lend themselves to combining data from different children to document group progress? Issues such as these should be considered carefully before an assessment plan is developed.

TYPES OF ASSESSMENT INSTRUMENTS

Norm-Referenced Assessment

An important goal of intervention programs is to accelerate the developmental progress of the children they serve. Showing increases in child IQs or MDIs is one way to document this acceleration. Norm-referenced instruments such as the Bayley Scales of Infant Development (Bayley, 1969) and the Cattell Infant Intelligence Scale (Cattell, 1960) are diagnostic instruments that identify those children who are developing normally, those who are developing slowly, and those who are developing at an accelerated rate. Norm-referenced instruments must be administered in a stardardized fashion by trained examiners using materials provided by the test developers. A child's performance is compared to that of other children in the normative sample. Each child's status is described in terms of deviation from the mean of children of comparable age (MDI or IQ), or in terms of the age at which normally developing children accomplish the same sort of tasks (developmental age or mental age). These tests, which have been widely used to assess child progress in intervention programs, are relatively reliable and the MDIs or IQs are presumed to reflect rate of development. A child who obtains a near 50 IQ on several longitudinal assessments is progressing at about half the rate of normally developing children. This child is maintaining his or her developmental rate. On the other hand, a child who obtains an IQ of 50 on one evaluation and an IQ of 60 on a subsequent evaluation is considered to be developing at an accelerated rate.

However, using infant tests as rate measures has been questioned by several authors because 1) infant development appears discontinuous across developmental stages (McCall, 1979), 2) the kinds of skills included in the tests vary widely from one age level to another (Johnson, 1982), and 3) these tests are rarely sensitive to change in significantly handicapped children (sensory or motor impairments in these children preclude "normal" interactions with the test materials under standardized conditions).

Criterion-Referenced Assessment

Criterion-referenced instruments assess current skills, identify instructional objectives and, when used longitudinally, assess progress in meeting instructional objectives. In theory, a criterion-referenced test is developed by first breaking important skills into their component parts, then into the separate steps necessary for mastering each part. A test item is developed to match each step, specifying what will be learned as well as the criteria by which task mastery will be judged.

With criterion-referenced assessment, a child is described by the tasks that have been mastered, not be comparisons with other children. The tasks not yet mastered become instructional goals, beginning with the easiest tasks and progressing to the more difficult ones. Progress is documented by the number of new tasks mastered between assessment points.

For infants, criterion-referenced tests are problematic because of the difficulty in identifying appropriate test content. Little is known about

which infant skills are related to critical childhood skills (reading readiness, for example). In addition, little information exists regarding which infant skills are most amenable to training and which are determined primarily by neurological maturation. Most criterion-referenced instruments for infants have been developed not on the basis of a theory nor on the basis of identifying prerequisite skills for later adaptive behavior and academic achievement. Rather, these instruments have been developed by adopting items from various norm-referenced tests, then grouping these items into domains (according to item content), and arranging them according to the mean age at which normal children pass the items.

Assessing progress that results from intervention is difficult when using criterion-referenced tests. These instruments include items inappropriate for seriously handicapped infants; in addition, no standards exist to indicate how many items would have been mastered simply through maturation. To deal with this problem, many instruments assign developmental ages to the test items, then compute ratio scores (developmental age divided by chronological age, multiplied by 100) for each domain. If the ratio increases during the intervention period, then it is assumed that progress results from intervention. This assessment procedure contains many methodological problems (Johnson, 1982), among these: 1) developmental ages assigned to each test item represent the mean age at which normally developing children passed the item, but do not reflect the variability of children in mastering the skill; 2) the steps between each item reflect unequal intervals of difficulty; and 3) the items have been drawn from a variety of developmental tests representing different normative samples.

One way of resolving these methodological problems is to dispense with the notion of developmental ratios (White, 1979), and instead, compute the percentage of test items passed in each domain at each assessment point. Percentages for children who are receiving intervention and those who are not can then be compared. Or, these percentages can be compared between domains targeted and those not targeted during a particular intervention. Percentages in all domains can be compared between programs providing different forms of intervention. Despite the problems with this procedure, it remains a sound method of documenting progress due to intervention and is worthy of further consideration.

Ordinal Assessment

Ordinal assessment refers to instruments that arrange the component skills of long-range objectives according to degree of difficulty. Existing infant tests that claim to be ordinal tests are based on Piagetian theory (e.g., Uzgiris & Hunt, 1975). Piaget's theory defines seven domains of cognitive development, as well as six developmental stages that occur within each domain during the sensorimotor period (birth to approximately two years of age). The assessment includes items that will identify the stage of a child's development within each domain.

Although not intended as instructional guides, ordinal tests for infants have been used as criterion-referenced tests in infant education. Since these tests focus primarily on cognitive and language development, it is necessary

to choose additional instruments for assessing the broader objectives of most intervention programs. Although final goals and the steps for reaching them are better defined, problems still arise in documenting the progress related to intervention. As with criterion-referenced tests, it is difficult to determine how much progress is due to maturation and how much to training. In addition, the cognitive items contain heavy sensory and motor requirements, making it difficult to assess the progress of many handicapped infants.

Nontraditional Assessment

Nontraditional assessment is any procedure, formal or informal, which gives information that may be unavailable from published norm-referenced or criterion-referenced tests. A renewed interest has emerged recently in finding reliable ways to assess cognitive skills in children whose physical handicaps or behavioral patterns prevent use of traditional assessment procedures. Among the nontraditional procedures are: 1) a procedure for assessing discrimination and memory in children functioning under six months of age (Fagan, 1975); 2) a procedure for determining whether a child can learn a sequence of events and can recognize an alteration of that sequence (Kagan, Kearsley, & Zelazo, 1978); 3) a systematic presentation of items that tend to elicit smiling or laughter at different ages in infants (Cicchetti & Sroufe, 1976), and 4) a variety of operant learning procedures that assess a child's ability to learn how to produce an interesting environmental event (e.g., Brinker, 1981; Goldman, 1980).

These nontraditional procedures correlate with other, more traditional measures of cognitive development while making very limited demands on the child's motor competency. Nontraditional instruments are able to identify cognitive competence not always evident in other assessment procedures. In non-traditional assessment, child progress may be observed by repeated measures; however, it would be difficult to demonstrate that progress is due to intervention in any but the operant learning procedure.

THE MULTIVARIATE ASSESSMENT APPROACH

Simeonsson, Huntington, and Parse (1980) have argued for a broader child assessment than that which is offered by the assessment procedures normally used in intervention programs. They suggest using a variety of instruments to assess child progress since the instruments sensitive to changes in one handicapped child may not be sensitive to changes in another child. In addition to traditional measures of development they suggest including measures of adaptive behavior and temperament in tests that assess the progress of significantly handicapped infants.

Adaptive Behavior

The definition of adaptive behavior poses interesting problems in the area of infancy. Traditional measures (e.g., the Vineland Social Maturities

Scale, Doll, 1965; the Preschool Attainment Record, Doll, 1966) include few test items for the infancy period; also, these items emphasize motoric behaviors. Little emphasis is placed on the nonmotoric components of social and communicative behaviors often considered important components of adaptation in infancy. The new revisions of the Vineland Social Maturities Scale (i.e., the Vineland Adaptive Behavior Scales, Sparrow, Balla & Cicchetti, 1984), provides more items for the infancy period and includes four domains: communication, socialization, daily living skills, and motor skills. The Vineland Adaptive Behavior Scales may provide a better measure of adaptive behavior in the infancy period than has been available previously. However, no reports of its use with very young children have yet been published.

Another promising instrument is the Carolina Record of Individual Behavior (Simeonsson, Huntington, Short, & Ware, 1982). The CRIB, which expands the Infant Behavior Record used with the Bayley Infant Scales (Bayley, 1969), is sensitive to changes in severely handicapped children for whom the usual developmental measures are insensitive. The CRIB is also responsive to changes in children who have specific sensory and motor impairments (Simmeonsson et al., 1982).

Temperament Measures

Most temperament measures used with infants (e.g., Revised Infant Temperament Questionnaire, Carey & McDevitt, 1978; Infant Behavior Questionnaire, Rothbart, 1981) depend on the judgements of parents or other caretakers regarding the characteristics of the infants. Because of this, temperament measures are criticized as being more a measure of the adults that fill them out, rather than an objective measure of infant characteristics. Such measures may be seriously considered as measures of child progress. However, if one adopts a transactional view of development, then such measures may be considered for assessing child progress. In the transactional view, it is assumed that the development of family members is affected by their interactions with each other. If a parent reports changes in a child's temperament, it may be assumed that changes have occurred in the family system which are important in the child's development. Several possibilities exist: the child may have changed, thus changing parent perception; the child actually may have changed little, but altered parental perceptions may create changes in the child; or both parent and child may be changing in a mutually reinforcing way. In any event, temperament measures do reflect child progress.

An Appropriate Test Battery

When choosing a multivariate assessment approach, the purposes for assessing progress as well as the characteristics of the children to be assessed must be considered carefully. If a program is serving a heterogeneous population of handicapped youngsters, the best strategy for documenting individual child progress may be to tailor assessments individually for each child, choosing instruments appropriate to their abilities and disabilities. Once this choice is made, however, it will be impossible to use your group data with traditional research designs to demonstrate program effectiveness.

Another strategy for documenting progress is to choose one comprehensive criterion-referenced instrument for use with all the children. This instrument would be one or more instruments sensitive to behavioral and temperamental characteristics in all the children. Finally, a variety of other instruments tailored to meet the needs of individual children would be chosen. For example, one might choose the Vulpe Assessment Battery (Vulpe, 1977), the Hawaii Early Learning Profile (Furuno, O'Reilly, Hosaka, Inatsuka, Allman, & Zeisloft, 1979), or the Assessment Log of the Carolina Curriculum for Handicapped Children (Johnson-Martin, et al., in press); these could be supplemented with the CRIB (Simeonsson et al., 1982) and the Revised Infant Temperament Questionnaire, (Carey & McDevitt, 1978). Then, some nontraditional techniques could be used to assess cognition in the more severely motor-impaired children, along with the Reynell-Zinkin (1979) for use with children who are visually impaired. In addition, the Parenting Stress Index (Abindin, 1983) might be used to determine how much stress the infants exert on their families. (From a transactional viewpoint, a decrease in family stress would be one additional indicator of child progress.)

WHAT PROCEDURES SHOULD BE USED TO DETERMINE IF PROGRESS HAS BEEN MADE?

The standard scientific method for demonstrating the effectiveness of any treatment is to assign subjects randomly to either a treatment group and a control group, or to a group that receives treatment A and a group that receives treatment B. Pre- and posttesting will determine whether the groups have made different degrees of progress or change resulting from the treatments. Obviously, this design is problematic in the area of early intervention. Ethical considerations prohibit assigning children to a control or no-treatment group. Some programs have used existing conditions to by-pass the ethical problems, e.g., by assessing the progress of children on a waiting list and comparing it to the progress of children already in a treatment program (this is not random assignment; and the comparability of the groups is a concern). Although a "treatment A vs treatment B" design may be more ethically sound, difficulties do exist in identifying different treatments and in ensuring the comparability of the groups. Some studies compare the progress of children in an intervention program with children whose development is reported elsewhere in the literature (for example, Hanson's, 1977, comparison of the children in her Down's syndrome intervention program with data available on the development of home-reared Down's syndrome children not in early intervention programs). While appealing, such comparisons are fraught with methodological problems related to comparability of the groups.

In attempting to document intervention effectiveness, many prgrams have used either a standardized or a criterion-referenced test to which developmental levels were attached for demonstrating developmental rate changes within individuals and within the group.

Alternatives to Experimental and Control Group Designs

New methods of documenting intervention effectiveness must be found if early intervention programs are to receive continued support. A number of

intervention programs are considering some viable alternatives to group design. These alternatives are: 1) multiple baseline design, 2) a single-subject design comparing the longitudinal development of a child in intervention with a normative sample, and 3) goal-attainment scaling designs.

Multiple Baseline Designs

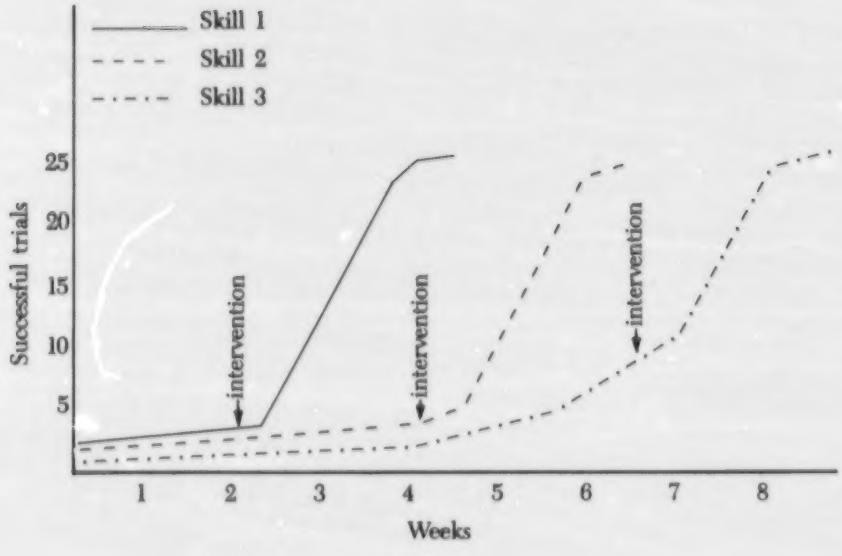
Developed to assess behavioral changes, multiple baseline designs involve measuring several behaviors at intervals until their stability under natural conditions is ascertained. Then, intervention is introduced for one of the behaviors while all other behaviors continue to be measured at regular intervals. Once the success criterion is obtained for the first behavior, intervention is started for a second behavior; again, all other behaviors continue to be measured. This procedure is repeated for each of the remaining behaviors. The baseline of the behaviors is the standard against which the effects of intervention are judged (Tawney & Gast, 1984) -- see Figure 1.

The multiple baseline procedure assumes that target behaviors are functionally independent -- a difficult assumption to make in the area of infant development. For example, one cannot be sure that progress in head control is independent of progress in making sounds or progress in grasp patterns. How-

Figure 1

Multiple Baseline Graph

(hypothetical results)



ever, variations of the multiple baseline procedure have been used in some infant programs to document progress due to intervention. (e.g., Barerra, Routh, Parr, Johnson, Arendshorst, Goolsby, & Schroeder, 1976; Johnson-Martin et al., in press) and should be considered further.

The multiple baseline procedure can be applied to an individual child by assessing his or her ability to perform various skills over several days or weeks. Instruction is then started for one of these skills, while data is collected regularly on mastery of the other skills. After the one skill targetted for intervention is mastered, intervention is initiated for the next skill, and so forth. Charting progress over time should reveal changes in skill acquisition that coincide with intervention. Skills for which treatment is delayed serve as a baseline against which progress is measured in the skills targetted for intervention.

In a second modification of the multiple baseline procedure, domains of development are divided into two or more groups (preferably through random assignment). A child or group of children is assessed in all domains at regular intervals; intervention is initiated in one group and continued for several months. Then, intervention is discontinued in these domains (although assessment continues) and initiated in the second set of domains. Over time, progress in the domains targetted for intervention may be compared with progress made in domains not targetted for intervention. This procedure was used in the preliminary study by Barrerra et al. (1976) to document the intervention effectiveness in a small group of children; the field test procedures of the Carolina Curriculum for Handicapped Children (Johnson-Martin, et al, in press) also uses this procedure. In both instances, evidence was provided for statistically significant differences in progress made intervention areas.

Many interventionists argue against delaying intervention in one area of development in order to use the delayed area as a baseline to assess another area. But early childhood special educators must realize that no intervention program can achieve everything that needs to be done for a child at one time. Further, it is unnecessary to delay all intervention in, for example, the gross-motor area while working in the language area. Each of the traditional domains of development can be divided into smaller domains which can then be used as the basis of the multiple baseline procedure.

In field testing the Carolina Curriculum, for example, 24 skill areas were identified, and children were assessed in each of these areas. For three months, intervention was carried out in all five motor areas and in half of the other areas, with an equal representation of strong and weak areas for each child. At the end of three months each child was reassessed in all 24 areas. Intervention for the next three-month period continued in the motor areas and in the areas where no intervention had been made in the first three months. Then, progress (i.e., the total number of items passed) was compared between those nonmotor areas worked on and those not worked on during each three-month period (Johnson-Martin et al., in press).

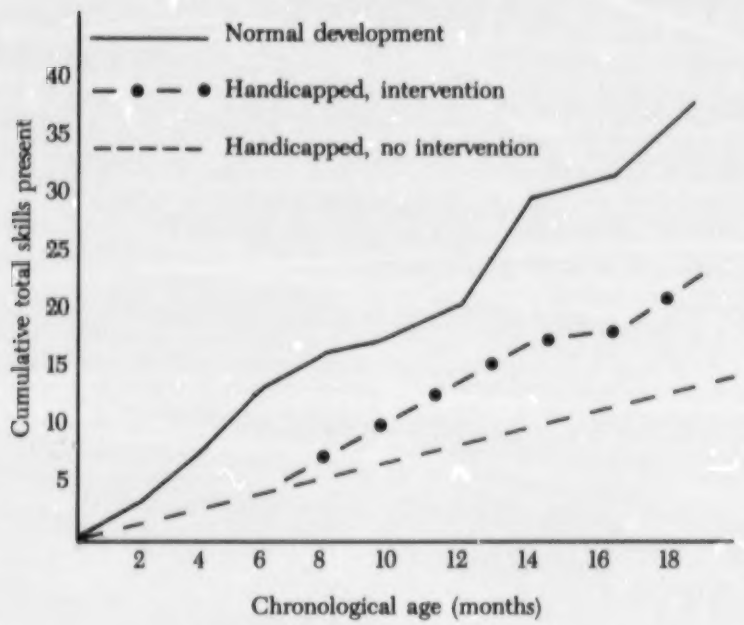
Comparison to normative skill acquisition. This alternative to group design, suggested by Sommers, McGregor, Lesh, and Reed (1980), is a procedure that considers the cumulative total of skills and behaviors present (measured on any preferred assessment devise). This total is plotted on a graph against

expected attainment for the normative sample over a longitudinal period. Pre-intervention assessment indicates the number of skills a handicapped child has mastered. These are plotted on a graph to show the comparison to normal children at a given time. After intervention is initiated, periodic assessments continue; this data is also plotted on the graph. A change in the slope of the curve presumably reflects changes in development due to intervention. Data plotted for a hypothetical case is presented in Figure 2.

Figure 2

Comparison of Intervention with Normal Development and Expected Development without Intervention

(hypothetical results)



Goal Attainment Scaling. A procedure developed by Kiresuk and Lund (1976), Goal Attainment Scaling involves setting goals for individuals in a number of domains and assessing the degree to which the goals are achieved. A formula permits the derivation of a Goal Attainment Score (GAS) for each individual. GASs may be compared regardless of the assessment instruments or other bases upon which the goals were developed. Likewise, a program can be evaluated by examining the average GAS of the clients. This procedure allows an analysis of group data without the constraint of having to use a common assessment instrument for every individual in the group. The Goal Attainment Scaling method has been widely used with adults in business and in human services. It has also been used in a few studies of children in mental health

and school settings (Simeonsson, Huntington, & Short, 1982). According to Simeonsson et al. (1982), a number of features make Goal Attainment Scaling a useful evaluation strategy for handicapped infants and children: it is directly relevant to therapeutic concerns and can use unique child measures; it can make use of the goals and objectives that are already a part of most intervention programs; it allows for differential weighting of goals; and the values derived from the use of GASs are both numerically and conceptually meaningful. The most essential problem with Goal Attainment Scaling is determining the basis for selection of the goals for each child. Theoretically, intervention produces development or change beyond that expected from maturation. Such goal setting requires astute judgments on the part of the interventionist. One program might look effective and another ineffective if the former program set unrealistically low goals and the latter set goals that were unrealistically high. To prevent this kind of bias, interventionists from other programs should be invited in to evaluate the children independently and compare the goals that have been set and attained with those for similar children in their own programs or in other programs with which they are familiar. At this stage of our knowledge, the consensus of experienced interventionists is probably the best measure that is available for judging the validity of the goals selected for the Goal Attainment Scaling procedure. As an additional safeguard for some types of clients, progress measured by Goal Attainment Scaling can be compared with progress suggested by norm-referenced assessments.

Child Progress or Program Effectiveness

The developmental progress of children that presumably results from intervention is often the primary, if not the only, measure used to document program effectiveness. This is unfortunate since early intervention programs provide many services that may not specifically affect developmental progress, i.e., the mastery of developmental milestones. Programs that involve parents often have a significant effect on the way these parents feel about themselves and their children, on their ability to serve as effective advocates for their children, and on their ability to interact effectively with their children.

Changes in parental attitudes, coupled with child-directed program efforts, may create important behavioral changes in children -- changes not reflected in developmental measures. For example, a nonresponsive child may become responsive, an irritable child may become calmer, while a passive child may grow more active, and so forth. Surely, such changes represent child progress and should not be overlooked when documenting the effectiveness of a program. Developmental measures are only one factor determining the long-term outcome of early intervention effectiveness (Lazar & Darlington, 1982).

REFERENCES

Abidin, R. R. (1983). Parenting Stress Index. Charlottesville, Virginia: Pediatric Psychology Press.

Barrera, M. E. C., Routh, D. K., Parr, C. A., Johnson, N. M., Arendshorst, D. S., Goolsby, E. L., & Schroeder, S. R. (1976). Early Intervention with Biologically Handicapped Infants and Young Children: A Preliminary Study with Each Child as His Own Control. In T. D. Tjossem (Ed.), Intervention Strategies for High Risk Infants and Young Children. Baltimore: University Park Press.

Brinker, R. (1981, April). Patterns of Learning by Handicapped Infants. Paper presented at the biennial meeting of the Society for Research in Child Development, Boston.

Carey, W. B., & McDevitt, S. C. (1978). Stability and Change in Individual Temperament Diagnoses from Infancy to Early Childhood. Journal of Child Psychiatry, 17, 331-337.

Cattell, P. (1960). The Cattell Infant Intelligence Scale. New York City: The Psychological Corporation.

Cicchetti, D., & Sroufe, A. (1976). The Relationship Between Affective and Cognitive Development in Down's Syndrome Infants. Child Development, 47, 920-929.

Doll, E. A. (1966). Vineland Social Maturity Scale. Circle Pines, Minnesota: American Guidance Service.

Doll, E. A. (1965). Preschool Atainment Record. Circle Pines, Minnesota: American Guidance Service.

Fagan, J. F. (1975). Infant Recognition Memory as a Prsent and Future Index of Cognitive Abilities. In N. R. Ellis (Ed.), Aberrant Development in Infancy: Human and Animal Studies. Hillsdale, New Jersey: Erlbaum.

Furuno, E., O'Reilly, K. A., Hosaka, C. M., Inatsuka, T. T., Allman, T. L., & Zeisloft, B. (1979). Hawaii Early Learning Profile. Palo Alto, California: Vort Corporation.

Goldman, B. D. (1980). Transfer of Contingency Awareness in Six-Month-Old Infants. Unpublished doctoral dissertation, University of Waterloo, Waterloo, Ontario, Canada.

Hanson, M. (1978). Teaching Your Down's Syndrome Child: A Guide for Parents. Baltimore: University Park Press.

Johnson, N. (1982). Assessment Paradigms and Atypical Infants: An Interventionist's Perspective. In D. Bricker (Ed.), The Application of Research Findings to Intervention with At-Risk and Handicapped Infants. Baltimore: University Park Press.

Johnson-Martin, N. M., Jens, K. G., Attermeier, S. M. (in press). Carolina Curriculum for Handicapped Infants (and infants at risk). Baltimore: Paul Brookes Publishing Company.

Kagan, J., Kearsley, R., & Zelazo, P. (1978). Infancy: Its Place in Human Development. Cambridge, Massachusetts: Harvard University Press.

Lazar, I., & Darlington, R. (1982). Lasting Effects of Early Education: A Report from the Consortium for Longitudinal Studies. Monographs of the Society for Research in Child Development, 47(2-3), Serial No. 195.

McCall, R. B. (1979). The Development of Intellectual Functioning in Infancy and the Prediction of Later I.Q. In J. D. Osofsky (Ed.), Handbook of Infant Development (pp. 707-741). New York City: John Wiley and Sons.

Reynell, J., & Zinkin, D. (1979). Reynell-Zinkin Scales: Developmental Scales for Young Visually Handicapped Children. Windsor, England: NFER Publishing Company, Ltd.

Robthbart, M. K. (1981). Measurement of Temperament in Infancy. Child Development, 52, 569-578.

Simeonsson, R., Huntington, G. S., & Parse, S. A. (1980). Expanding the Developmental Assessment of Young Handicapped Children. In J. J. Gallagher, (Ed.), New Directions for Exceptional Children: Young Exceptional Children, San Francisco: Jossey-Bass.

Simeonsson, R., Huntington, G. S., & Short, R. J. (1982). Individual Differences and Goals: An Approach to the Evaluation of Child Progress. Topics in Early Childhood Special Education, 1(4), 71-80.

Simeonsson, R., Huntington, G. S., & Short, R. J., & Ware, W. G. (1982). The Carolina Record of Individual Behavior: Characteristics of Handicapped Infants and Children. Topics in Early Childhood Special Education, 2(2), 43-55.

Somers, D., McGregor, G., Lesh, E., Reed, S. (1980, December). A Rapid Method for Describing the Efficacy of Early Intervention Programs for Developmentally Disabled Children. Mental Retardation, 275-278.

Sparrow, S. S., Balla, D. A., & Cichetti, D. V. (1984). Vineland Adaptive Behavior Scales. Circle Pines, Minnesota: American Guidance Service, Inc.

Tawney, J. D., & Gast, D. L. (1984). Single-Subject Research in Special Education. Columbus, Ohio: Charles E. Merrill Publishing Company.

Uzgiris, I. C., & Hunt, J. McV. (1975). Assessment in Infancy: Ordinal Scales of Psychological Development. Urbana: University of Illinois Press.

Vulpe, S. G. (1977). Vulpe Asessment Battery. Toronto, Ontario, Canada: National Institute of Mental Retardation.

White, O. R., & Haring, N. G. (1979). Evaluating Educational Programs Serving the Severely and Profoundly Handicapped. Unpublished manuscript.

Chapter 4
Social Validity

by Phillip Strain

WHAT IS CHILD PROGRESS?

Early childhood special education professionals who attend to the issue of child progress usually concentrate on the psychometric properties of child outcome measures (e.g., reliability, predictive validity). Unfortunately, social validity or social significance of outcome indices is too often ignored.

Examining the social validity of an outcome addresses this basic question: Does the outcome directly or indirectly help the child and family meet society's demands (Wolf, 1978)? This broad question of social validity can be further divided:

- Following intervention, do the child and family function more like their peers in the social community?
- Do members of the child and family's peer group perceive that the intervention effects are positive, and do they behave more positively toward the clients after intervention?

Seldom are either of these two questions addressed in early intervention outcome research. Yet, ignoring these social validity concerns may set the stage for "wash-out" of intervention effects, minimal public support for intervention programs, and the perpetuation of statistically significant but socially trivial outcomes.

Let us take, for example, two tried-and-true outcome indices in early intervention programs -- children's cognitive functioning and parents' teaching skills -- and weigh them against the social validity criterion.

Social Validity Vs. Measurement of Cognitive Functioning

Changes in children's cognitive functioning, as assessed by standardized IQ instruments or developmental scales, are often used as an outcome in intervention programs for developmentally delayed children. Is change in cognitive functioning a socially valid index of intervention impact?

Regretably, the answer is unclear. For handicapped populations within a relatively narrow band of performance (mild mental retardation to above average intelligence), IQ is a good predictor of socially accepted and valued performance, such as: timely progress through school, academic achievement near or at grade-level, gainful employment, and positive self-evaluations. For moderately and more severely handicapped groups, the relationship between IQ level and social adaptation is uncertain. IQ has little relevance to socially valid criteria such as less restrictive educational environments and deinstitutionalization. These opportunities tend to be based not upon IQ but upon the individual's social skills (lack of deviant, antisocial behaviors). Further, there is no evidence at this time to suggest that a certain level of IQ increase is even perceptable to significant social agents. Most intervention efforts that use IQ as the outcome have shown gains of one standard deviation or less. It is doubtful that the expanded repertoire necessary to produce a 10-, 15-, or 20-point IQ increase will have any significant impact on the way the child functions in society.

Social Validity Vs. Measurement of Parents' Teaching Skills

Parent's teaching skill is a frequent target in early intervention programs. More often than not, some form of criterion-referenced assessment is used to assess teaching skills prior to instruction and after instruction. A wealth of literature is now available that points to the efficiency of many training models and the impact of parents' newly acquired skills on their children's behavior. One might argue that social validity is certain when positive changes in child behaviors can be linked directly to parents' improved teaching skill. But, there is scant evidence that teaching parents to teach has a positive, facilitative influence on the parent, the parent-child relationship, or the family. We do not know, for example, if parents who receive certain training are therefore more satisfied and fulfilled in their lives; if they now receive more support from family and friends; if they are more willing to persist with child instruction; if they have better marital relations; if they are more active in pursuing their life goals; if they perceive that the overall quality of their life has changed; and so on. Of course, it is possible also that negative outcomes, as yet unmeasured, may accrue from parent training.

By no means is the goal here to criticize the use of IQ measures, parents' teaching skills, or any other outcome index. Rather, the intent is to draw attention to the many socially significant outcomes of early intervention and to suggest that they be considered in the area of assessment of child progress. Widespread use of a child outcome does not, ergo, secure social validity.

The examples below offer several suggestions for expanding our professional repertoire of socially valid indices of assessing child and parent outcomes of early intervention.

Socially Valid Child Outcomes

Most early intervention programs do an excellent job of improving children's accuracy on various tasks or developing behavioral competencies that did not exist prior to intervention. Few programs, however, have examined the fluency with which children perform their newly acquired skills. Yet, responding at a satisfactory pace is a major determinant of the social acceptability or validity of many skills. An accurate but belabored signature may go unnoticed in the preschool or kindergarten, but it is certainly a skill deficit in the early elementary grades (Kerr & Lambert, 1982). Similarly, accurate counting of objects 1 to 10 is seldom made time dependent in our early intervention efforts, yet, fluency in numerosity is absolutely essential for the most rudimentary of math operations (Strain & Kerr, 1981). Finally, many programs measure the number of words that a child learns to say in a "word-calling" context and spend a considerable amount of time teaching new words. However, reading experts generally agree that an acceptable rate of response along with good comprehension are the hallmarks of successful readers.

In addition to including measures of fluency, early intervention programs can also improve the social validity of child outcomes by selecting levels of target behavior (e.g., percent of time spent in positive social interaction, percent of time working independent of adult supervision) that are associated with successful performance in future educational and social settings. Using both teacher-completed checklists/ questionnaires and direct observation of successful children, a number of investigators (e.g., Strain, in press; Vincent et al., 1980; Walker & Hops, 1976) have pinpointed levels of academic and social performance that characterize competent children in preschool, kindergarten, and the early elementary grades.

Socially Valid Parent Outcomes

To date, parent outcomes in early intervention programs have been limited primarily to a narrow range of behaviors that occur within the mother-child dyad. Consistent with the educational emphasis of most early intervention programs, teaching behaviors exhibited by parents are the predominant target of treatment to parents and focus of parent evaluation.

Though these teaching skills may be important to the child's long-term progress (e.g., Lovaas, 1981; Strain, Steele, Ellis, & Timm, 1982), many indirect parent outcomes of early intervention remain undetected. For example, many parents with handicapped children become quite insular; that is, they do not tend to go on outings, have extended friendships, or generally share their positive and negative experiences with others. In many ways, their behavior is structurally similar to several primary characteristics of clinical depression. Recent work by Wahler and Dumas (in press), documents repreated incidence of reported depression by parents of disabled children. Family insularity and depressive characteristics are both powerful indicators of present and future adjustment problems. Assessment and amelioration of these problems could greatly add to the social validity of early intervention outcomes.

HOW CAN CHILD PROGRESS BE MEASURED?

Indeed, there are many advantages to using standardized tests to document the efficacy of an early intervention program. For example, results can be readily compared with other intervention programs using the same measures, the psychometric properties of the tests are known, and uniform procedures for administration and scoring reduce the chance of spurious findings. On occasion, however, the very procedures for test administration and scoring can bias outcomes against detecting a significant intervention effect.

When we use most standardized tests we must accept two basic assumptions: the children or parents are sufficiently motivated -- by the testing material and despite limited examiner feedback -- to do their best; and the response mode(s), the content, and the way the test items are presented do not unduly discriminate against the testee. For many handicapped young children, these assumptions do not conform to our best information about the fragile nature of human behavior and the reactivity of assessment. So, when we use standardized tests to measure progress of handicapped children, we get results based on possibly false assumptions.

Motivation

Over a half-century of research has been conducted on the effects of motivation on test performance. The overwhelming amount of this research indicates that children attempt more test items and produce significantly higher test scores when they are reinforced for "good test-taking behavior." Some might argue that the logical approach is simply to reinforce preschool handicapped children during test-taking, but such an approach would render the intervention evaluation invalid. When standardized tests are used, it is essential that the test assumptions be met to the maximum extent possible (Maheady, Sainato, & Maitland, 1983).

A major issue for the early intervention researcher, then, is to identify the extent to which the problems of nonmotivated assessment may affect any particular evaluation. For many groups of preschool handicapped children, the bias created by nonmotivated testing is likely to be substantial. For example, several categories of handicapped children (e.g., autistic, severely and profoundly mentally retarded) have as a major defining characteristic the inability to acquire new skills without carefully programmed use of motivational incentives or reinforcement. Indeed, extensive research has been devoted to uncovering methods for identifying the idiosyncratic motivational variables necessary for learning with severely disabled groups (e.g., Carr, 1977).

The contrast between the heavy use of motivational incentives in an intervention program and the absence of motivational incentives in the assessment situation suggests that children's test performance would be less than optimal. Where children receive extensive feedback for their daily performance in intervention programs, nonmotivated assessment may be particularly punishing and biased.

Mode, Content, and Presentation

Many standardized tests result in underestimations of children's performance because the response mode, content, and presentation of the assessment items discriminate against handicapped preschoolers. The majority of standardized tests require the testee to respond with an expressive language response or a pointing response. For many preschool handicapped children, expressive language production is under very tenuous control. That is, their verbal behavior may vary greatly with particular instructors, settings, and stimulus materials. For other preschool handicapped children, their verbal repertoire may be quite extensive and not situation-specific; yet, their articulation is such that examiners who are not intimately familiar with certain verbal omissions or substitutions may unreliably score many responses.

Standardized assessment instruments presuppose that testees have some exposure or experience with the particular assessment items. As Duncan et al. (1981) suggest, the contents of many standardized tests (particularly those aimed at cognitive assessment) assume that children have been exposed to: traditional family living arrangements; household items; animals; certain toys; and so on. This assumption is likely false for children from disadvantaged backgrounds, children who reside in other than traditional homes and children whose handicaps have limited their access to environmental events common to nonhandicapped children.

Standardized tests are intended to be presented in ways that will ensure that children understand what response is required but will not prompt the correct answer. However, more often than not, the examiner gives the child a verbal command to engage in some behavior. This presentation format assumes that testees are generally compliant and that the command (e.g., "Show me _____." "Point to _____." "What is this?") is understood. We know, of course, that many preschool children are in treatment programs because of their oppositional behavior and general noncompliance to adults' requests. For many severely disabled children, verbal comprehension is extremely restricted. There are countless children, for example, who can make many correct discriminations when asked, "Show me _____." If these same children are asked, "Point to _____," they may not respond at all. It is a regrettable irony that the limited, carefully routinized cues often needed to teach severely disabled children may limit performance on many standardized tests.

Summary

The tendency when using standardized tests is to interpret nonresponding and incorrect responding as valid evidence of a skill deficiency. Among the general population of preschool handicapped children there are children who:

- know many answers but are not motivated to respond;
- cannot respond correctly because they simply have no relevant experience with the test item;

- could likely score much higher on a standardized test if the response mode of the test conformed to their learning characteristics;
- do not respond generally to any requests made by adults
- would likely get many more items correct if the stimulus items were presented in a familiar format.

For all of these children and many others, evaluation outcomes based on standardized tests may limit the opportunity to detect real intervention effects. Measurement of specific intervention effects using techniques of single-subject designs currently offers the best alternative to standardized testing.

WHAT DESIGNS OR PROCEDURES SHOULD BE USED TO DETERMINE PROGRESS?

In recent years, there has been a rapid increase in the use of single-subject designs to evaluate the effects of early intervention programs. To understand the true advantages and limitations of these designs, it is important to deal initially with a few of the more onerous problems.

First of all, single-subject designs (perhaps a more descriptive term might be, "subjects-as-their-own-control designs") that appear in the current literature often include multiple subjects in numbers equal to those in experimental or control groups in the early education efficacy literature. Second, the use of these designs is not limited to the evaluation of behaviorally oriented intervention projects. Single-subject designs have been used to study the effects of health safety regulations, toothbrushing, anti-litter campaigns, racial awareness programs, "I"-message use, and countless school-related interventions. There are no single-subject design requirements that automatically exclude the evaluation of even the most strident nonbehavioral intervention. Finally, there is the issue of external validity, or the confidence one can have in generalizing the findings from one outcome study to members of the same or a similar subject group.

It has been argued that single-subject designs do not offer any evidence that suggests findings can be generalized. In truth, single-subject and group designs offer two very different solutions to the important external validity problem. Group designs rely on the random or matched assignment of subjects to experimental groups and the laws of probability to determine the likelihood that an outcome could have been produced by chance. The lower the "chance occurrence" of the findings the more confidence one can have that similar subjects will be affected in a like manner if the intervention were repeated. Single-subject designs, on the other hand, attempt to show (usually with a more limited number of subjects) that the intervention in fact will yield similar effects. Chance is reduced to zero. In addition, individual single-subject designs are viewed as one in a series of replication needed to validate any intervention.

These replications involve the same intervention implemented by different change agents with different target subjects.

There are a number of characteristics of single-subject designs that make them particularly attractive for early intervention outcome research.

Continuous Behavior Monitoring

Single-subject designs require that the intervention targets (e.g., improved preacademic performance, greater social participation, less tantrums, no episodes of self-injury) be monitored on a continuous, day-to-day basis. Though a labor-intensive requirement, continuous monitoring yields vital information not available in group designs. Specifically, the timing and pattern of intervention effects are preserved. A pre-post, group design method can show that an intervention produced a certain effect, but the continuous monitoring in a single-subject design can pinpoint when the effect took place, if it was gradual or abrupt, and if level at posttest was the optimum change produced.

Precise Description of the Independent Variable (i.e., Intervention)

Individual single-subject experiments are considered to be one of a series of replications. To facilitate replication, the independent variable must be defined and measured with such precision that the intervention can be reproduced with high fidelity. This requirement is of great benefit to program developers, and it is the only way to ensure an exact replication. Therefore, in descriptions of independent variables in single-subject research, the following requirements should be met: instructional behaviors of teachers, parents, etc., are clearly specified; these instructional behaviors are measured systematically as to their appropriate applications; independent observer reliability confirms the accuracy of the measurement of the instructional-behaviors; materials used in instruction are specified, and their use is monitored reliably; and the setting (i.e., physical arrangements, ratio of adults to children or handicapped to non-handicapped, etc.) for instruction is specified and reliably monitored. When any of these requirements is missing, the relationship between intervention and outcome is clouded, and scientific replication becomes a hit-or-miss proposition. It should be emphasized that scientific replication means something quite different than institutional adoption of a treatment approach or service delivery model, though the later is often called replication. Scientific replication, which Sidman (1960) argues to be the very essence of demonstrating that the presence of variable X produced Y outcomes, demands that treatments be so well documented that another individual could create the same instructional context and, it is hoped, produce the same outcomes with similar children (Strain, 1981).

By contrast, intervention variable descriptors in group designs often (but not always) are limited to nominal presentations such as, "the Ameliorative Curriculum was used," "developmental milestones were employed," or "teacher reinforcement was given." Obviously, an exact replication could not follow on the heels of such an inadequate description.

Public Policy and Best Practice

One of the hallmarks of P.L. 94-142 and a basic tenent of best educational practice is the individualization of assessment and treatment. Practitioners from all related persuasions seem to agree that the benefits gained by careful individualized programming are worth the added costs; indeed, an impressive data base supports this notion. Single-subject designs in which an individual's own beginning performance is the primary standard against which the intervention effect is compared is the closest we have come to eliminating the dysfunctional separation between good practice and research of scientific integrity.

I do not consider that single-subject designs have inherent disadvantages vis a vis assessing early intervention outcomes. I do believe, however, that consumers and potential users should be aware of the following caveats.

Professional Myopia

A modest, albeit occasionally powerful, minority in the scientific community look upon single-subject designs as something less than scientific and their users as something worse than Satanic. The existence of this minority is not a reason to avoid single-subject designs, but it may be worthwhile to analyze the biases of your intended audience before a final decision is made about a treatment outcome design.

Discontinuous Dependent Measures

Some potential child outcomes are not appropriately measured by, or compatible with, continuous measurement. Included are measurements of IQ, achievement, and motor development that rely on testing procedures that are confounded by repeated use over a short period of time.

Design Fluctuations

Most procedural details of between-group research can be specified from the outset. However, in single-subject designs, the incoming data dictates decisions such as: length of a baseline or preintervention assessment phase; when to change experimental phases; and adding another intervention component. The necessary flexibility built into single-subject designs demands close, usually daily monitoring.

Interactions Between Client and Intervention

Often it is important to know if an intervention is more or less effective as a function of such child characteristics as: level of intelligence, age, handicapping condition, or past treatment history. Currently, no acceptable single-subject designs can address intervention-

client interactions. Factorial designs should clearly be chosen in such instances.

REFERENCES

Carr, E. G. (1977). The Motivation of Self-Injurious Behavior: A Review of Some Hypotheses. Psychological Bulletin, 84, 800-816.

Duncan, D., Sbardellati, E., Maheady, L., & Sainato, D. (1981). Nondiscriminatory Assessment of Severely Handicapped Individuals. TABH Journal, 6, 17-22.

Kerr, M. M., & Lambert, D. L. (1982). Behavior Modification of Children's Written Language. In M. Hersen, P. M. Miller & R. Eisler (Eds.), Progress in behavior modification (Vol. 13). New York City: Academic Press.

Lovaas, O. I. (1982, August). The Young Autism Project. Paper presented at the American Psychological Association, Washington, D.C.

Maheady, L., Sainato, D., & Maitland, G. (1983). Motivated Assessment: The Effects of Extrinsic Rewards on the Individually Administered Reading Test Performance of Low, Average, and High IQ Students. Education and Treatment of Children, 6, 37-46.

Sidman, M. (1960). Tactics of Scientific Research. New York City: Basic Books.

Strain, P. S. (1981). Conceptual and Methodological Issues in Efficacy Research with Behaviorally Disordered Children. Journal of the Division of Early Childhood, 4, 111-124.

Strain, P. S. (in press). Identification of Peer Social Skills for Preschool Mentally Retarded Children in Mainstreamed Classes. Applied Research in Mental Retardation.

Strain, P. S., Steele, P., Ellis, T., & Timm, M. A. (1982). Long-Term Effects of Oppositional Child Treatment with Mothers as Therapists and Therapist Trainers. Journal of Applied Behavior Analysis, 15, 163-169.

Vincent, L. J., Salsbury, C., Walter, G., Brown, P., Gruenewald, L. J., & Powers, M. (1980). Program Evaluation and Curriculum Development in Early Childhood/Special Education: Criteria of the Next Environment. In W. Sailor, B. Wilcox & L. Brown (Eds.), Methods of Instruction for Severely Handicapped Students. Baltimore: Paul H. Brookes.

Walker, H. M., & Hops, H. (1976). Use of Normative Peer Data as a Standard for Evaluating Classroom Treatment Effects. Journal of Applied Behavior Analysis, 9, 156-168.

Wolf, M. M. (1978). Social Validity: The Case for Subjective Measurement or How Applied Behavior Analysis Is Finding Its Heart. Journal of Applied Behavior Analysis, 11, 203-214.